

MEASUREMENT

TEXT-BOOK SERIES

EDITED BY PAUL MONROE, PH.D.

TEXT-BOOK IN THE HISTORY OF EDUCATION.

By PAUL MONROE, PH.D., Professor of History of Education, Teachers College, Columbia University.

SOURCE BOOK IN THE HISTORY OF EDUCATION.

FOR THE GREEK AND ROMAN PERIOD

By PAUL MONROE, PH.D.

PRINCIPLES OF SECONDARY EDUCATION.

By PAUL MONROE, PH.D.

TEXT-BOOK IN THE PRINCIPLES OF EDUCATION.

By ERNEST R. HENDERSON, PH.D., Professor of Education and Philosophy, Adelphi College.

DEMOCRACY AND EDUCATION. An INTRODUCTION TO THE PHILOSOPHY OF EDUCATION.

By JOHN DEWEY, PH.D., Professor of Philosophy, Columbia University.

STATE AND COUNTY SCHOOL ADMINISTRATION.

SOURCE BOOK

By ELLWOOD P. CUBBERLEY, PH.D., Professor of Education, Stanford University, and EDWARD C. ELLIOTT, PH.D., Professor of Education, University of Wisconsin.

STATE AND COUNTY EDUCATIONAL REORGANIZATION.

By ELLWOOD P. CUBBERLEY, PH.D.

THE PRINCIPLES OF SCIENCE TEACHING.

By GEORGE R. TWISS, B.Sc., Professor of the Principles and Practice of Education, Ohio State University.

THE PRUSSIAN ELEMENTARY SCHOOLS.

By THOMAS ALEXANDER, PH.D., Professor of Education, Teachers College, Columbia University.

A HISTORY OF MARRIAGE AND THE FAMILY, REV. ED.

By WILLYSTINE GOODSELL, PH.D., Associate Professor of Education, Teachers College, Columbia University.

THE HISTORY OF THE EDUCATION OF WOMEN.

By WILLYSTINE GOODSELL, PH.D.

STATISTICAL METHOD.

By TRUMAN L. KELLEY, PH.D., Professor of Education, Harvard University.

FOUNDATIONS OF EDUCATIONAL SOCIOLOGY, REV. ED.

By CHARLES C. PETERS, PH.D., Professor of Education in Ohio Wesleyan University.

SOURCE BOOK IN THE PHILOSOPHY OF EDUCATION, REV. ED.

By WILLIAM H. KILPATRICK, PH.D., Professor of Education, Teachers College, Columbia University.

MEASUREMENT, REV. ED.

By WILLIAM A. MCCALL, PH.D., Professor of Education, Teachers College, Columbia University.

MEASUREMENT

By

WILLIAM A. MCCALL, PH.D.

Professor of Education

Teachers College, Columbia University

A Revision of *How to Measure in Education*

THE MACMILLAN COMPANY

New York

COPYRIGHT, 1939,
By THE MACMILLAN COMPANY

ALL RIGHTS RESERVED—NO PART OF THIS BOOK MAY BE
REPRODUCED IN ANY FORM WITHOUT PERMISSION IN WRITING
FROM THE PUBLISHER, EXCEPT BY A REVIEWER WHO WISHES
TO QUOTE BRIEF PASSAGES IN CONNECTION WITH A REVIEW
WRITTEN FOR INCLUSION IN MAGAZINE OR NEWSPAPER

• Printed in the United States of America •

Published February, 1939
Second Printing February, 1940

To
My Students

Especially to those who prefer to live on the *frontier*, to keep in contact with the new without losing touch with the old, to be acutely sensitive to the future without being unaware of the past, and to work *toward* the millennium while working *for* it, believing that they will arrive sooner if they seize every opportunity to make an advance however small instead of advocating everything or nothing thus leaving to the ultra-conservatives the practical control of the educational process.

PREFACE

There are tides in the affairs of men—moving inexorably. Today as in few periods of history, an old order changes yielding place to a new, vaguely perceived. The working classes, having developed leadership, are rising. Democracy, heretofore confined largely to politics, is, under mass pressure, being injected into economics. The brotherhood of man is coming—perforce.

Men and women, groping blindly, ask the schools for light—ask the schools to acquaint their students with matters more urgent than Redskins, windmills and wooden shoes, igloos, Latin, and dodoes. They beg teachers to inform the new generation, through realistic experience, with the fundamentals of finance and economics so they may think more clearly than their parents about these issues and not be a prey to every plausible propagandist.

There is an answering stir among teachers. They feel this world ferment. They perceive how pitifully incompetent their former pupils are to grapple in a virile manner with the emerging problems. They humbly admit their own ignorance. A few are thrilling their students with the challenge of significant activities, resisting the temptation to leave life alone and stay within the sanctuary of well-thumbed books.

Twenty years ago, I found measurement in danger of becoming, like medicine, the property of a professional élite. I wrote and invented that its mysteries might be banished and it might be made available to all teachers. More and more in late years I have felt that measurement is in danger of becoming an ally of the *status quo*. It has tended to spin endlessly about itself a protective web of statistical intricacy, untroubled by any philosophical spark, and undisturbed by the world's travail. This book has been written not only to preserve the good that is in the *status quo* but also to yank measurement out of its statistical complacency, infuse in it a new spirit, sensitize it to the life that is outside as well as inside textbooks, place it, as in former years, in the van of education.

I confess that the book fails to achieve a perfect integration

of my science and my philosophy. The world is in transition. Education is in transition. I too am in transition.

I confess another fault, hoping thereby to win charity. The book is, in part, a compilation of my publications during several years—publications naturally devoted to my own researches and inventions. When I see these together in one volume, I am embarrassed by the frequency with which my own productions are used for purposes of illustration. “Of making many books there is no end,” says the Bible. Of the making of a book like this, there is no end. Since I can ill afford the enormous amount of time and energy required to correct this fault, I solicit, dear reader, your charity, not being enough of a pure scientist to accept with equanimity your condemnation.

WM. A. MCCALL

December, 1938

CONTENTS

BOOK ONE

PLACE OF MEASUREMENT IN EDUCATION

CHAPTER	PAGE
I. A PHILOSOPHY OF MEASUREMENT	3

BOOK TWO

CRITERIA FOR THE SELECTION AND CONSTRUCTION OF STANDARD AND TEACHER-MADE TESTS

II. HOW TO SELECT AND CONSTRUCT TESTS—VALIDITY	29
III. HOW TO SELECT AND CONSTRUCT TESTS—RELIABILITY AND OBJECTIVITY	55
IV. HOW TO SELECT AND CONSTRUCT TESTS—NORMS AND SCALES	61
V. HOW TO SELECT AND CONSTRUCT TESTS—SCORING	66
VI. HOW TO SELECT AND CONSTRUCT TESTS—INSTRUC- TIONS	80
VII. COMPREHENSIVE LIST OF TESTS AND TEST PUBLISHERS	91

BOOK THREE

USE OF STANDARD TESTS FOR GROUPING PUPILS

VIII. HOW TO ADMINISTER TESTS AND OBTAIN AND INTER- PRET GRADE OR AGE SCORES	137
IX. HOW TO COMBINE GRADE OR AGE SCORES	146
X. HOW TO PREPARE CLASS RECORD SHEETS	150
XI. HOW TO CLASSIFY PUPILS	156
XII. HOW TO CLASSIFY WHEN PROMOTIONS ARE SEMI- ANNUAL	199
XIII. HOW TO HANDLE SPECIAL SITUATIONS	207

BOOK FOUR

PROGRAM OF MEASUREMENT FOR PROGRESSIVE
SCHOOLS

CHAPTER	PAGE
XIV. THE COMPREHENSIVE TESTS	215
XV. INSTRUCTIONS FOR USING THE COMPREHENSIVE TESTS	278
XVI. HEALTH, DYNAMIC, PERSONALITY, AND MATERIALS TESTS	298

BOOK FIVE

GUIDANCE AND EVALUATION OF TEACHING BY
MEASUREMENT

XVII. SUBJECTIVE MEASUREMENT OF THE TEACHING PROC- ESS	321
XVIII. OBJECTIVE MEASUREMENT OF THE TEACHING PROCESS	345
XIX. OBJECTIVE MEASUREMENT OF THE EFFECTS OF THE TEACHING PROCESS	354
XX. TESTS AS TEACHING INSTRUMENTS	367
XXI. DIAGNOSTIC MEASUREMENT	383
XXII. EFFICIENCY OF PUPILS, TEACHERS, PRINCIPALS, AND SUPERINTENDENT	402

BOOK SIX

SCHOOL MARKS AND REPORTS

XXIII. VARIETIES OF MARKING SYSTEMS	419
XXIV. CRITICAL EVALUATIONS OF MARKING SYSTEMS . . .	423
XXV. PREPARING TO OPERATE THE GRADE SCORE MARKING SYSTEM	430
XXVI. MARKING EXAMINATION PAPERS IN TERMS OF GRADE SCORES	435
XXVII. DIAGNOSTIC INTERPRETATION OF GRADE SCORE MARKS	440
XXVIII. RECORDS AND REPORTS TO PARENTS	445
XXIX. CLASSIFICATION, PROMOTION, AND GRADUATION . .	452
XXX. SOME QUESTIONS AND ANSWERS	462

BOOK SEVEN

PRESENTATION OF TEST RESULTS

CHAPTER	PAGE
XXXI. GRAPHIC METHODS	473

BOOK EIGHT

HOW TO SCALE TESTS AND COMPUTE STATISTICAL
MEASURES

XXXII. REFERENCE POINTS AND SCALE UNITS	495
XXXIII. SCALES AND THEIR CONSTRUCTION	501
XXXIV. CONSTRUCTION OF SCALED SCORING INSTRUMENTS	513
XXXV. STATISTICAL METHODS	519
INDEX	531

LIST OF TABLES

TABLE	PAGE
1. A Substantially Complete List of All Primary, Elementary, High School, and College Tests and Publishers of Tests	92
2. G Table for Thorndike-McCall Reading Scales	143
3. G Table for Woody-McCall Mixed Fundamentals in Arithmetic Scales	144
4. Class Record Sheet for School Having Annual Promotion	151
5. Calculation of Differences between Gp's and Norms	175
6. Selection of Classification Table	176
7. Table for a School Which Attempts to Do 0.9 Standard Grade per Year, Showing the Automatic Classification of Pupils into Grades on the Basis of Any G (Grade) Score	177
8. Table for a School Which Attempts to Do 1.0 Standard Grade per Year, Showing the Automatic Classification of Pupils into Grades on the Basis of Any G (Grade) Score	180
9. Table for a School Which Attempts to Do 1.1 Standard Grades per Year, Showing the Automatic Classification of Pupils into Grades on the Basis of Any G (Grade) Score	183
10. Classification Standards for a School Using the 0.9 Classification Table	188
11. Classification Standards for a School Using the 1.0 Classification Table	189
12. Classification Standards for a School Using the 1.1 Classification Table	190
13. Special Cases in the Sample School in Which the Conservative Classification Was Not Followed	192
14. Distribution of Changes Made in Reclassifying School X by Means of Educational Tests	194
15. Distribution of Changes Made in Reclassifying School Y by Means of Educational Tests	195
16. What Happened to the Specially Promoted Pupils of School X	196
17. Class Record Sheets for School Having Semi-Annual Promotion	200
18. Computation of Gt, Grade 1L	203
19. Calculation of Differences between Gp's and Norms	204

TABLE	PAGE
20. Indices of Reliability for the McCall Intelligence Test, Educational Background Questionnaire, and Comprehensive Achievement Test	289
21. Grade Scores, Age Scores, Norms, Lists of Answers and Record Sheet for <i>A Comprehensive Test Program</i>	290
22. Requirements for Communicable Diseases	300
23. The Author's Personality Quotient	315
24. Relation of B Scores and F Scores	356
25. Relative Merits of Three Marking Systems	427
26. Sample Work Sheet—Annual Promotion	431
27. Amounts to Be Added to G Scores to Obtain Projected Gi Scores	432
28. Sample Work Sheet—Semi-Annual Promotion	432
29. Sample Page from the Teachers Record Book	433
30. Assignment of Marks on Arithmetic Examination.	436
31. Assignment of Marks When Several Pupils Have the Same Number Right	437
32. Showing Probable Errors and Significant Differences between Gi and Subject G Scores for Examinations of Various Lengths	442
33. Sample of Report Card to Parents	445
34. Sample of Pupils Cumulative Record Card for Elementary Schools.	448
35. Conversion of Grade Score into Letter Marks	449
36. Sample Conversion Table for Grade 5	449
37. Summary Sheet for Grade 5	453
38. Calculation of Differences between Gp's and Norms	455
39. Selection of Classification Table	456
40. 0.9 Classification Standard Table	456
41. 1.0 Classification Standard Table	457
42. 1.1 Classification Standard Table	457
43. Showing the Need for Equal Units of Measurement	501
44. G Table for Haggerty Reading Examination Sigma 1	504
45. Showing How to Scale Total Scores	505
46. Showing S.D. Distances Corresponding to a Given per Cent	507
47. Showing How to Widen the Range of a T Scale	508
48. Showing How to Determine Age Norms for a Test	511

TABLE	PAGE
49. Showing P.E. Distances Corresponding to Each per Cent .	514
50. Determining P.E. Distances in Merit between Composition Specimens	515
51. Showing How to Compute a Coefficient of Correlation When Data Have Been Tabulated in a Contingency Table . . .	523
52. Showing the Coefficients of Correlation between Attendance and Six Hypothetical Causes	525
53. Showing Decreases in Error of Prediction with Increases in r	526

LIST OF DIAGRAMS

FIGURE	PAGE
1. Overlapping of Educational Ages for Two Adjoining Grades	193
2-29. Illustrating Standard Methods for Graphic Presentation by Means of Bar and Curve Diagrams	475-481
30. A Sector Diagram	484
31. A Sectioned-Bar Diagram	485
32. A Combination Bar-and-Sectioned Diagram	486
33. Frequency Surface Showing Identification of Components	488
34. Graphic Method of Constructing G Table	502
35. Per Cent of Attendance	521

BOOK ONE



PLACE OF MEASUREMENT IN EDUCATION

CHAPTER I

A PHILOSOPHY OF MEASUREMENT

THESIS 1.¹ THE ULTIMATE TEST OF ALL THINGS IS THE HAPPINESS THEY YIELD

Even the more thoughtful of human beings seldom appear to have found any "anchor for a drifting world" or to have very clearly defined any ultimate goal toward which to shape their action. Yet it would seem altogether impossible to construct an intelligent plan of individual action or a scheme of education until some such fundamental objective has been formulated and accepted. The formulation of an ultimate educational objective for oneself is not, as has been frequently supposed, a subject solely for the amusement of speculative philosophers.

All are substantially agreed that the school is merely one of many agencies for facilitating and improving the social process. Hence the inquiry comes down to this: What is this whole social process for? Unfortunately this simplification does not make the answer evident.

When we think philosophically, we trust to the reasonableness of our ideas to gain them acceptance. In imagination we perch upon some lofty eminence whence we can see in perspective, with our mind's eye, the wave of civilization sweeping from the Orient westward, and we attempt to predict the future and guess the ultimate objectives of the human race.

Neither the ultimate objectives formulated by the theologians nor those stated by most traditional philosophers will offer much guidance in discovering educational goals. The focus of the typical theologian is at a point too far beyond this mundane sphere, and the typical philosopher has written as though a man were a cold-blooded, unemotional, thinking mechanism designed only for grinding out speculations concerning the purely intellectual aspects of the cosmos. Many of the philoso-

¹ Most of Theses 1 and 2 is quoted from *You and College* by McCall, Balch, and Herring with the permission of Harcourt Brace and Co., New York City.

phers of the past have been in bondage to the hypothesis of logic that we think clearest when we see the problem in perspective. They have gone away from human beings in order to look down at them. They have followed Kepler's advice and tried to think God's thoughts after Him. They have assumed that the Deity knows whither we are going, or they have tried to wrest the secret from inscrutable Nature.

If, instead of a theological and philosophical approach, we make a philosophical-psychological approach, and if, instead of trying to make known the hypothetical purposes of an Unknown, we inquire about the purposes of each man, woman, and child, we get an answer which has profound significance for religion, sociology, and education and which is eminently practical for our purposes. Will many persons attest that their most fundamental purpose is to increase Complexity, or evolve Mind, or prepare for citizenship in Valhalla? Not at all. They may consider these among their many purposes, but hardly more than that. What is this social process for? The behavior and thought of every individual in the world is witness to the truth of this conclusion: *The social process is to satisfy human wants, desires, or purposes—to achieve happiness.*

The social process may be a means of attaining some far-off objective. To man it is a means of securing for himself maximum happiness by satisfying the wants bred in him by nature and instilled in him by environment. To assume that an acceptance of these ideas means that man's education will be founded upon low desires or selfish interests is to be convicted of an ignorance of modern psychology and of the nature of human purposes. The traditional distinction between selfishness and unselfishness is untenable in the light of modern psychology. Both the selfish and the unselfish man act in accordance with their strongest desires. The difference is that the so-called selfish man gets his satisfaction from realizing wants which cross the wants of others while the unselfish man gets his happiness from satisfying purposes of his which happily coincide with the purposes of others. Not all wants are low or narrow. Many of man's purposes reach into future generations and even toward a future life. Religious teachers and social reformers have intense purposes which would certainly not be called either low or narrow. There is a limitless range in the quality of human wants. Some men

have brought their wants to such a high plane that they would be glad to forego certain of their present purposes to aid all human beings.

Despite the fact that the same criterion is imbedded in everybody's mind, men differ about it. Some say it is duty; some, pleasure; some, happiness; some, adjustment to circumstances or fate. Some say it is the welfare, in a defined sense, of the individual, and others, the welfare of the state or of society. Still others, of course, hold different views. The problem of judging the values of human activities cannot be well handled without facing and answering, as we have done, the question; *What is the highest good?* The analysis of the problem and the evaluation of its various answers are the subject of books and courses in ethics. This is not the place to argue the question; and yet it has somehow to be answered, for the problem is one of the greatest, and, like all great problems, is not supplied with fixed, ready-made rules that we can adopt as a whole, understand immediately, and follow in a routine fashion. We shall have to choose between studying values and shifting along without ever really facing and solving the problem.

In this book *happiness*, as previously suggested, will be treated as the answer to the question: What is the greatest good? What is it that everybody ought, so far as he can, to seek for all men, including himself? Those who do not agree with the answer may think of it as an illustration, and should then of course use their own answer instead.

Since happiness will mean different things to different people, in spite of the interpretation already presented, it is right to clear away at the outset certain additional possible misunderstandings. The goal, happiness, is broad, taking in all mankind. It means happiness for every man, woman, and child in the world. No exception ought to be admitted or in practice allowed whenever it can be prevented. It does not mean happiness for one or a few at the expense of the rest, or for one nation against the others, or for one generation in preference to its elders or juniors. When exceptions cannot be avoided, these ought to be treated as fairly as possible. Often it is wise and necessary for one to forego a happiness in order to bring happiness to others.

Happiness is of many kinds. It includes sensory pleasures like the enjoyment of warmth, food, color, form, and sex; it includes

also delight in friends, books, music, contemplation, work, sports, and every other thing that is enjoyed. It is sometimes wise to forego happiness of one sort in order to insure greater happiness of other sorts.

Happiness is for the whole length of life, not just for the moment. It is frequently wise to forego happiness of the moment in order to provide a more lasting happiness in the future. But since life is made up only of a series of presents, we ought to make sure we do not habitually sacrifice the present. We should learn to live happily in the present in such a way that the future, when it comes, will be a series of happy present moments.

The goal, happiness, is broad, taking in all mankind; it is deep and high, comprising all levels of human enjoyment; and it is long, including both the present and the future. Three misconceptions, then, of the meaning of happiness, and consequently of its use and effects as a goal, are here answered. They may be suggested by the three phrases: not *mine* merely; not *pleasures* merely; not *now* merely.

Happiness, we have said, is the chief criterion. No matter what we are judging, we can test it by its contribution to happiness. Everything anybody does, ought to be chosen and guided so as to lead to this end. Eating is good when we enjoy it, or when it leads to nourishment and health and therefore to happiness; but it is bad when we do not enjoy it, or when we have already eaten as much as we can enjoy or as much as will lead through health to happiness, or when we have a fever and eating more will bring misery. It is good to bring together in co-operative pursuits two groups like Negroes and whites, but only when it will lead to happiness for more people. If it can lead only to unresolvable conflict and therefore away from happiness for many people, it is to that extent bad.

Since it is a conventional view that we should act always for the greatest good of the greatest number or for the greatest fairly shared good of all, and hence that we should make decisions for our own lives as if we were a disinterested jury, we may be tempted to accept this view uncritically. Before we commit ourselves irrevocably to this doctrine, we should reflect upon the following rather disturbing considerations, some of which, though vital to life's most important decisions, have often received scant attention in the history of philosophy:

1. The happiness of others will receive much emphasis through our affections, our desire to make persons in general happy, and our fear of others' resentment.

2. We know ourselves well enough to make decisions fairly satisfactorily for ourselves, whereas we cannot feel very accurately for others.

3. Much of the world's unhappiness is caused by parents' or associates' making decisions for those whose nature they do not understand and whose future they cannot predict.

4. On many vital matters most persons conceal their real preferences and even seek to convey the opposite impression, thus adding to the difficulty of one person's making estimates for another.

5. Many Mr. and Mrs. Grundys make a business of claiming interest in the decisions of others and of resisting decisions which the next generation will applaud.

6. Though the conventional view has been held for centuries, few persons if any live in accordance with it. This fact of itself casts grave doubts upon its validity.

7. The Creator of man made him essentially egocentric, perhaps for good and sufficient reasons.

8. Many eminent thinkers favor the greatest good of the best portion of the population only, excluding sometimes the intellectually deficient, and sometimes other groups regarded as inferior.

9. It may be argued with much reason that the surest way to secure the greatest happiness for all is for each to aim only at his own happiness.

10. One of America's most distinguished philosophers argued for the good life as against the happy life, not clearly realizing that the measure of the goodness of an act is the total quantity of happiness resulting from it in the long run.

11. One of the latest books on philosophy argues that in judging the worth of an act, we must consider the quality as well as the quantity of the resulting happiness. The author of that book has failed to perceive that the amount of happiness yielded in the long run is the index of its quality (though not of course of the nature of the feeling).

12. The distinction between total happiness of the population, and the average happiness is too seldom applied; and yet

major decisions, such as whether or not to approve Japan's invasion of Manchuria, Italy's attack on Ethiopia, or programs for the limitation of population may turn on this distinction. It is possible to increase the total amount of happiness of a nation by increasing the population while at the same time making everybody a little more miserable.

When so many significant aspects of life's most vital question have been fallaciously conceived or inadequately considered, even at times by the most distinguished thinkers, we are fully justified in thinking to our own conclusion as though the world were still young—as it is—and we were its first adequate philosophers—as we may be.

The author and others are now engaged in extensive experimentation with various methods for measuring an individual's present status in happiness. The *Comprehensive Achievement Test*,¹ described in a later chapter, includes a test of happiness. It is just a matter of time until prophetic tests will be invented. None of the controversies in education or life can be fully settled until an adequate happiness test is available.

THESIS 2. IT IS PROPER FOR MOST TESTS TO MEASURE SECONDARY TRAITS

Happiness is a criterion which it is always proper to use. All the others, like speed, nourishment, and coöperation, are only parts of a larger picture which lose meaning when they become separated from the whole. Thus speeding may kill people, nourishment may injure a sick man, and coöperation may sometimes cause distress.

Yet these secondary criteria are useful in suggesting ways in which happiness may be both judged and sought. They may help you to estimate the value of doing something and so to decide what to do. If we set out to make happiness secure through communal activities, we may get guidance for action from such criteria as: participate often; choose important events; work efficiently; take the lead; be resourceful in suggestion; share fairly with other people; and try to have others do likewise. We shall hardly prepare a successful meal if we think of nothing but happiness while we are planning it; we must

¹ Published by Laidlaw Brothers, Chicago, Ill.

also think of what foods our guests or family like, of food prices we can safely afford to pay, of the values of foods for health and therefore for happiness.

These two ways of estimating the values of life are represented by two kinds of questions:

First Kind of Question

If I do this, will more happiness result?

If I do this, will happiness come to more people?

If I do this, will happiness be more permanent?

Second Kind of Question

If I do this, will it lead through coöperation to happiness?

If I do this, will it lead through nourishment to health, and through health to happiness?

The second kind of question is better than the first because it includes the first and adds something useful to it. We should not think of deciding whether to go to college just by saying it would cause more happiness to do so. The problem is too complex for that. We should think of a number of differences or consequences involved: money, respect, understanding of the world, a more satisfactory marriage, greater likelihood of contributing to science or art or industry. But so tricky are these secondary criteria that we could be richer, more respected, more learned, better married, more able and more useful to science and still create more misery than happiness in the world. We could use money, respectability, learning, marriage, skills, and contributions to science or industry in such a way as to subject other men to our will and thereby make a few happy and many miserable.

There is, then, a

Dangerous Kind of Question

If I do this, will it make for more success?

If I do this, shall I become wealthy?

If I say this, will I be telling the truth?

Such questions are too isolated to be safe. If we use them by themselves, we or others are sure to suffer in the long run. Any of the lesser criteria, taken by themselves, may give us a distorted world. We shall do better to use happiness along with success, friendliness, truthfulness, and other secondary criteria.

Every day we witness flagrant misjudgments brought about

by using some lesser criterion as if it were the chief one. We can observe everywhere the abuse of such standards as the profit motive, truth-telling, the welfare of institutions, and the understanding of the world as the chief end of college life.

The profit motive is used by many persons as the sole basis for judging their careers. Content with their own happiness or at most with that of their families and friends, they fail to inquire about the effects of their careers upon other people. In consequence, other people go hungry, have no chance to educate their children, are deprived of their essential right to the pursuit of happiness. Those who say the profit motive must go, think of it as a lesser criterion which ought to be tested by being put under the chief one. When they do test profit as a means to the happiness of all they find it wanting. To be justified, they think mere money-making would have to be proved an effective means to the successful pursuit of happiness for all.

Truth-telling is another subordinate criterion often wrongly raised to first rank. Some people appear to believe that they should always tell the truth. But others think that truth-telling, also, should be put in its place as a means to happiness. They think it would be wrong to tell a sick man of his exact condition when it would lessen his chances of recovery, or to inform people that they are stupid or their dinners dull and insipid, even if they are, when nothing is to be gained thereby for anybody's happiness.

The welfare of institutions is a third example. It is right to work, through the support of a school or club or state, for human happiness; but no institution ought to be encouraged which, on the whole and in the long run, makes for human misery or even for less happiness than would result from applying the support elsewhere. Institutional practices, too, even of only a few months' standing, have a way of getting absurdly magnified into "sacred traditions" and permitted to enslave and stultify both officers and members. In the act of restoring such reactionary societies to their full usefulness, when they are not entirely beyond reclaiming, revolutions are sometimes necessary which give pain to a few persons in order to bring happiness to many.

We have tests of ability in reading, writing, arithmetic, handling disagreements, choosing wisely and many other such

secondary criteria because we believe that the possession of certain abilities, attitudes, and ways of acting contribute to happiness. An inspection of all the forms of the *Comprehensive Achievement Test* and the *School Practices Questionnaire*¹ will disclose a particularly inclusive list of secondary criteria.

THESIS 3. THE ALLEGED CONFLICT BETWEEN MEASUREMENT AND GESTALT PSYCHOLOGY IS EQUIVALENT TO THE CONFLICT BETWEEN SECONDARY CRITERIA AND THE ULTIMATE CRITERION

Just as the ultimate criterion is somewhat more than a summation of secondary criteria, so the *gestalt* or whole child is somewhat more than the addition of his scores on an utterly comprehensive battery of intelligence, personality, and achievement tests.

But it is unwise to focus attention so fully on the uniqueness of individual organism that one tends to forget the more substantial similarities not only from person to person but also from generation to generation. To focus unduly on the unique aspects causes one to reject science and its concepts of reasonable prediction, to deny the possibilities of preparing any teaching plans or materials in advance and hence to deny any very practical help to all teachers except the most superior ones, and finally, in its most virulent form, to pass into a sort of do-nothing mysticism.

Certain more extreme exponents of this organismic view contend not only that any organism is more than the sum of its parts, but also that adding test scores is like trying to make a man by sticking together a head, a trunk, two arms, and two legs. A reading score cannot be properly compared to one leg. It is not a broken-off fragment of the mind. In a very real sense, a reading score tends to measure the entire organism functioning in that reading situation.

THESIS 4. MEASUREMENT IS ESSENTIAL TO THE MAINTENANCE AND INCREASE OF EACH GENERATION'S CAPACITY TO LEARN

There is now substantial acceptance by practically all persons who have dispassionately studied the evidence:

¹ Published by Laidlaw Brothers, Chicago, Ill.

1. That the quantity of the population in the Occident is steadily declining.

2. That the intellectual quality of this population is steadily declining.

3. That the best germ plasm is disappearing from the western nations at a rapid rate.

4. That present tendencies unless consciously interrupted will continue until involution has run its tragic course, destroyed our urbanized civilization, and evacuated city dwellers who are no longer competent to maintain its elaborate interdependent functions to a simpler rural environment which fosters the evolution of intelligence.

But is there no way to magnify intelligence by manipulations of the environment? Is not nurture far more potent than nature? The last is a much debated question, even though there is little room for debate if nature and nurture are measured in comparable units. Offer any reader of this book a choice, for a son, between a boy who is at the 25th percentile in the total population in native intelligence and is to be subjected to a 75th percentile nurture and another boy who is at the 75th percentile in intelligence and is to have a 25th percentile nurture and the reader will unhesitatingly choose the latter, knowing for a certainty that at any defined age the latter will surpass the former. Common observation is sufficient to prove that nature is much more powerful than nurture.

The mentality of pupils develops rapidly at first and then more and more slowly until maximum intellectual development has been reached at about age 20. Thereafter he may learn more things but he will never learn them with greater ease. By making education more efficient from birth to around age 20 and by placing pupils in general environments more suitable for mental growth, we may make the average and slightly less capable child of this generation slightly better able to master new problems at age 20 than the average child of the preceding generation. But it is very, very difficult for nurture to make amends for nature.

Furthermore, competent students of intellectual inheritance agree:

5. That educating one generation does not add one iota to the next generation's inherited capacity to learn.

Since most persons are not greatly interested in the fate of the race a thousand years hence, it might be pointed out that the social and economic chaos so characteristic of our own time must be regarded as probably permanent features of society. It must be evident to everyone that we shall never have a stable and contented society until we have real brotherhood of men and a genuine democracy in every aspect of life. It must be equally evident that a genuine democracy is well nigh impossible so long as we have the present tremendous range in human worth.

Hence it is highly important that society make a realistic study of exactly why the abler persons do not reproduce, and then somehow, somehow see to it that the population is filled up from the top only and not from the bottom only, thereby gradually enlarging the capacity to learn and narrowing the range.

It is scarcely the province of this book to propose plans for accomplishing these results. It will be sufficient to point out that intelligence tests, new as they are, provide a means whereby the average intelligence of the next generation could be markedly raised, for they might be used to locate individuals likely to produce gifted offspring. It should not be overly difficult to get the State to provide adequate motivation for persons of high intelligence to have more children and for persons of lower intelligence to have fewer. There would be numerous errors in identifying promising parents, but fewer errors than are made by the present methods of identification.

Unfortunately any measure designed to improve quality would almost certainly reduce quantity, and, until all the world is under a single control, drastic reduction in "cannon fodder" or factory fodder invites invasion and extinction. So it looks as though there is nothing significant that can be done except to increase our knowledge of heredity against the day when a world order is established and fundamental problems of every kind can be attacked in a fundamental way.

THESIS 5. TESTS PERFORM A VITAL SERVICE TO GOVERNMENTS

The existence of a vast body of ignorant and illiterate voters is not only a disgrace but a serious menace to the nation. Of the more than 1,500,000 who were drafted into the United States

Army during the World War nearly 25 per cent were unable to read or write. According to the *Journal of the National Education Association* for October, 1922, the nation's illiterates over 21 years of age could have outvoted the states of Pennsylvania, Maine, Michigan, Alabama, and California in the 1920 presidential election. The richest and most influential nation of the world is among the more illiterate! In the state of New York the number of illiterates increased between 1910 and 1920.

For this reason and others, New York state passed a law denying the ballot to all new voters in the state who could not produce evidence of graduation from the elementary school or pass a literacy test in reading and writing prepared and administered by the educational authorities of the state.

These authorities delegated to a Literacy Commission, of which the author was a member, complete responsibility for preparing the literacy test and determining the passing point on the test. The test designed by the writer consisted of a simple selection dealing with elementary matters of civic importance followed by simple, common-sense questions to which the candidate wrote answers. His reading ability was tested by his power to comprehend the selection well enough to answer the questions based upon it. The test of his writing ability was the functional one, namely, whether he could express himself legibly enough to make his meaning clear to the scorer of the test. Many forms of this test were prepared as a precaution to prevent politicians from coaching those being tested.

Although the passing point was fixed at a literacy equivalent to that of typical pupils graduating from the fourth grade the test was failed, the first time it was administered, by about 20,000 of the 100,000 who took it. This naturally provoked the opposition of politicians accustomed to deliver the illiterate vote. They undertook to have the literacy law declared unconstitutional on the ground that the test measured intelligence as well as reading ability. They were right, of course, but the highest court of the state declared that the test measured literacy and nothing but literacy—which is one way of determining the validity of a test!

Literacy tests for voters is merely one of numerous types of tests that are used or might be used advantageously by governments.

THESIS 6. "WHATEVER EXISTS AT ALL, EXISTS IN
SOME AMOUNT"¹

Since all sane persons accept this thesis it needs no qualification, but a qualified thesis will suffice for our purpose, namely, whatever change the teacher makes in a pupil must be a change in an amount of something. We teachers will scarcely insist that our effort makes no change in amount. Even though such were the result of our effort it would not so much disprove the thesis but rather prove our own inefficiency.

There is an ever-dwindling group who strenuously oppose the practical implications of the above thesis. They claim to be interested in the emancipation of education from the quantitative idea. Their effort is directed toward the qualitative in education. According to them there is in every person a non-quantitative quality—a

. . . something far more deeply interfused,
Whose dwelling is the light of setting suns.

Did they truly "see into the life of things" they would realize that there is never a quantity which does not measure some quality, and never an existing quality that is non-quantitative.

THESIS 7. ANYTHING THAT EXISTS IN AMOUNT CAN
BE MEASURED

At least half a dozen scales now exist by which it would have been possible to measure the quality of the Handwriting on the Wall. Faust said:

What she reveals not to thy mental sight
Thou wilt not wrest from her with levers and with screws.

But science has enormously increased the subtlety of levers and screws, and our mental sight is obtuse compared to some of our present-day mental tests.

It is possible to measure, at least crudely, an individual's love of a sunset or appreciation of opera. Theoretically the thesis is sound but whether practically we shall ever possess sufficient ingenuity to discover all the things that exist in

¹ Thorndike, E. L. *The Seventeenth Year Book of the National Society for Study of Education*, Part II, p. 16; Public School Publishing Co., Bloomington, Ill.

amount and then measure them with any great accuracy, a question. All that is necessary to accept for the present is that all the abilities and virtues for which education is consciously striving can be measured and be measured better than they ever have been. The measurement of initiative, judgment of relative values, leadership, appreciation of good literature and the like is entirely possible. We already have a scientific scale for the measurement of poetic appreciation. The measurements may not be as exact as we might wish, but they would have value.

THESIS 8. MEASUREMENT IN EDUCATION IS IN GENERAL THE SAME AS MEASUREMENT IN THE PHYSICAL SCIENCES

The two types of measurement are fundamentally alike because both measure physical manifestation. Neither add ability, nor good intentions can be measured by plunging a thermometer into a pupil's spiritual medium, but they can be measuring his behavior and judging his inner condition therefrom. Unless the witness is a habitual liar, psychologists can with considerable success, determine by means of a breathing curve, when a witness is not telling the truth.

In a still invisible future it may be possible to secure "movie" of a pupil's mental machinery when in operation and thus secure the desired information but for the present it is necessary to measure the product produced and, if desired, in the inner condition of the pupil.

Measurement must frequently meet the objection of being too materialistic. Listen to Gilder in "The Poet's Protest."

O man with your rule and measure,
Your tests and analyses!
You may take your empty pleasure,
May kill the pine, if you please,
You may count the rings and the seasons,
May hold the sap to the sun,
You may guess at the ways and the reasons
Till your little day is done.

To parody Wagner in "The Better Way," one would think that it was the purpose to measure human worth by the ell, the value of a life by the number of its years, the painter's canvas by the yard, or the work of the poet by the pound or bushel. A student

writes: "Measurement should not be applied where spiritual factors and ideal values are involved." Those educators who protest most violently against any such measurement of the pupil are daily probing his mental activity by methods which are comparable to the surgical operations of bygone ages. They find themselves in a position of disapproving the lover who estimates his lady's affection by the radius of the pupil of her eye under standardized lighting, and of approving the scientific father who soothes the mother for his punishment of their infant by saying: "I am not slapping an innocent soul but spanking a physiological reaction."

THESIS 9. ALL MEASUREMENTS IN THE PHYSICAL SCIENCES ARE NOT PERFECT

Physical measurements are, in general, more exact than educational measurements but education has no monopoly upon imperfect tests. There are tests which are now the rule in physical sciences for which an expert in educational measurements would blush. The general superiority of physical measurements is not due to the fact that they are radically different in kind. Physical measurements are subject to practically all the errors which trouble educational measurement. It is not that they do not exist in the former, but that they usually exist in such small amounts that the average person fails to see them. They are large enough to be the despair of experts in the various sciences. Thorndike¹ has given us an excellent statement of this point:

Nobody need be disturbed at these unfavorable contrasts between measurements of educational products and measurements of mass, density, velocity, temperature, quantity of electricity, and the like. The zero of temperature was located only a few years ago, and the equality of the units of the temperature-scale rests upon rather intricate and subtle presuppositions. At least, I venture to assert that not one in four of, say, the judges of the Supreme Court, bishops of our churches, and governors of our states could tell clearly and adequately what these presuppositions are. Our measurements of educational products would not at present be entirely safe grounds on which to extol or condemn a system of teaching reading or arithmetic, but many of them are far superior to measurements whereby our courts of law decide that one trade-mark is an infringement on another.

¹ *Op. cit.* p. 18.

But the imperfections of educational measurements are, in general, far more glaring than the majority of those made in physics, chemistry, and like sciences. Some may have gotten the impression that standard tests are perfect instruments. This is far from the truth. They have numerous and decided limitations.

A common criticism of educational measurement is that the tests measure a narrow, limited segment of a pupil's totality. Physical measurements tend to be more handicapped in this respect than educational measurements. Most of their measurements, such as measurements of length, width, weight, and temperature are exceedingly narrow abstractions and they are exceedingly useful too. A totality test for a pupil would certainly be useful but if we possessed one we would proceed immediately to construct tests for the detailed measurement of pupil abilities. Scales for the measurement of composition are useful, but scales for the measurement of the elements which go to make up composition are also useful. Teachers not only teach children "all over"; they teach them in detail. If tests are to aid instruction effectively, there is as much need for them to measure in detail as in totality.

THESIS 10. MEASUREMENT IS INDISPENSABLE TO THE GROWTH OF SCIENTIFIC EDUCATION

Exact measurement has made possible the rapid progress in the natural sciences. It has been stated that the amount of soap used is an index of the civilization of a country. The exactness of measurement is a good index of the status of a science. Consider where science would be without its meter, gram, ampere, volt, ohm, watt, henry, and the like. More than anything else it has been the absence of exact measurement which has kept education from the rank of a science. This plea for the development of those instruments which will make possible the progress of education as a science is made with knowledge of a recent statement by a prominent educator: "I think it would be disastrous if education were reduced to an exact science."

Richards,¹ in his presidential address before the American Association for the Advancement of Science, said: "Plato

¹ "The Problem of Radioactive Lead"; *Science*, January 3, 1919.

recognized, long ago, in an often-quoted epigram, that when weights and measures are left out, little remains of any art. Modern science echoes this dictum in its insistence on quantitative data; science becomes more scientific as it becomes more exactly quantitative."

In fact, measurement and education are like the twin girls whose hair the mother of many children braided together. Neither of the twins could move unless they both moved together.

Foote restates the above quotation in a nutshell when he says: "The day of guesswork must give way to definite facts supported by undebatable evidence."

There are those who tremble lest the development of education as a science will squeeze out of life its emotions and delicate perceptions. As well fear that woman suffrage or the "female" in industry will destroy gallantry among men. The roots of these fine things go too deep into human nature. In an especially unhappy mood Amiel writes: "Philosophy will clip an angel's wings," and again, "Science is a lucid madness engaged in tabulating its own necessary hallucinations." The basic function of Science is to help us to attain our objectives in the quickest and most economical way, whether the objectives be material or spiritual. Science is frequently looked upon as materialistic chiefly because only those persons who seek material objectives have had the good sense to secure the aid of Science. Haeckel, who has just drawn the line under his life's work, must have had in mind the unnecessary inefficiency of idealism when he wrote: "No cosmic problem was ever solved or even advanced by that cerebral function we call emotion." For centuries education has been like an emotional dog chasing a frantic tail. We have had a long line of great educational thinkers from Plato through Pestalozzi, and Froebel . . . to Dewey and beyond. "The old order changeth, yielding place to new," but no one seems to know whether the old or the new is better. In fact, there is grave suspicion that we move in an orbit whose form is the circle. These educational leaders are not answering questions. They are asking questions which do not occur to others. They are proposing problems for experimentation. The final answer to every educational question, except one, must be left to the educational measurer and must await the development of education as a science.

THESIS 11. MEASUREMENT IN EDUCATION IS BROADER THAN EDUCATIONAL TESTS

This book is not entitled EDUCATIONAL TESTS because there are other methods of measuring pedagogical products. Some estimate the quality of instruction by investigating the material equipment of libraries and laboratories and classrooms, or by the academic or professional training of the teacher. Others measure instruction by observing the teacher's method and by forming an opinion on the basis of these observations. Others base their judgments upon detailed observations of the behavior of pupils. Still others test the pupils by means of examinations. This book attempts to discuss the basic principles of measurement which apply not only to educational tests but to any sort of educational measurement. Some of the above methods of measuring educational results we are likely to have with us for some time to come. In our zeal for improving tests proper, we should not neglect the refinement of these methods. The main emphasis should be, and in this book will be, upon tests, because they offer the best promise for exact measurement. For if we can trust the experience in other fields, measurement by means of some sort of instruments will gradually replace all other forms. Finally, the book is not entitled EDUCATIONAL MEASUREMENT because education is deeply concerned with measurements which are not exactly products of instruction but about which educators need to be critical.

THESIS 12. TO THE EXTENT THAT THE PUPIL'S INITIAL ABILITIES OR CAPABILITIES ARE UNMEASURABLE A KNOWLEDGE OF HIM IS IMPOSSIBLE

A teacher needs the most intimate possible knowledge of a pupil in order to know what methods and materials to employ in order to help him most quickly to attain a desired goal. We partly know a pupil when we know the abilities and capabilities which he possesses. To determine the mere existence of an ability involves a crude measurement. But if we know no more than this we cannot tell whether a pupil has these abilities in sufficient quantities to permit him to matriculate for a Ph.D. in the university or just enough to enter the kindergarten. We

must know not only what qualities exist, but also in what amount they exist, and the more exactly we know this amount the better. Measurement is essential to a practical knowledge of psychology.

THESIS 13. "TO THE EXTENT THAT ANY GOAL OF
EDUCATION IS INTANGIBLE IT IS WORTHLESS"¹

We want to be able to answer at least three things about any goal: (1) What is the worth of the goal? (2) What is the location of the goal? (3) Is the pupil moving toward or from the goal? Measurement is necessary to answer each one of these absolutely vital questions. Suppose it be said that one goal of instruction is to produce in the pupil an ability to write. The worth of this goal depends upon an exact or crude measurement of how much penmanship contributes to the efficiency of a number of other superior activities. The goal has advanced little beyond perfect intangibility until it is located. How much ability to write? What speed? What quality? Even the worth of the goal cannot be answered until this location is made, since the worth varies with the quantity. The very words *how much* imply and in fact require measurement. Finally, it is necessary to answer the question: Is the pupil moving toward or from the goal? Without measurement the question is unanswerable.

THESIS 14. THE WORTH OF THE METHODS AND MATERIALS
OF INSTRUCTION IS UNKNOWN UNTIL THEIR EFFECT
IS MEASURED

The purpose of certain methods and materials is to help the pupil grow toward a certain goal. Do the methods employed accomplish their purpose? We cannot tell without employing measurement. For aught we know, the methods may be actually vicious. They may be forming habits which not only do not lead toward the goal, but which may be building up difficulties for another method by a subsequent teacher. It is equally true that the comparative worth of different methods and materials is unknown until their effect upon the pupil is measurable. This means that measurement is indispensable to the experimental selection of the most economical educational conditions.

¹ I am indebted to F. M. McMurry for this thesis.

Thus, measurement is everywhere in education and in our daily lives. Measurement is no rare freak. It gets up with us in the morning and goes to bed with us at night. The mile stone, the hand of the watch, the humble cup in the kitchen, the lengthening shadows of the trees on the grass, the spacing of the year into seasons, all indicate how ubiquitous measurement is. And measurement is just as immanent in the whole educational process as in life in general. There are other things in education besides measurement but they have no value so long as they are dissociated from it.

THESIS 15. MEASUREMENT OF ACHIEVEMENT SHOULD PRECEDE SUPERVISION OF TEACHING METHOD

Education is now being measured in two ways. When a child, I watched two coal miners lift a derailed car. Their efforts illustrate these two methods of measurement. A lever and fulcrum were brought, but the lever broke. A stronger lever was secured, but the fulcrum was too far from the car. Finally the proper adjustments were made and the car was lifted. Whether or not the car was lifted could be determined in two ways; (1) by measuring the length of the lever, the resistance of the fulcrum and the ground under the fulcrum, the weight of the men, the point of application of their weight, the distance of this point to the fulcrum, the distance from the fulcrum to the car, the weight of the car; or (2) simply by determining whether the car was actually lifted.

It is a fair assumption that the crucial purpose of elementary education is to make certain changes in children. To this end we have surrounded them with levers and fulcra in the shape of books, pictures, maps, tools, playthings, pedagogical methods, and with teachers who, with the pupils, will utilize these instruments as leverages to produce the desired changes.

Again it is a fair assumption that the schools should know whether their levers, fulcra, etc., are really producing the changes desired. As in the case of the derailed car, there are two methods of measuring these changes; (1) strength of lever, length of leverage, etc., become the number and nature of the books in the libraries, map facilities, blackboard space, and such, and the weight of the men becomes the number of diplomas

possessed by the teacher or else the amount of her skill in making provision for motive, initiative and such on the part of her pupils. (2) Whether the car is actually lifted is comparable to measuring directly the changes in the pupils.

Doubtless our relatively primitive ancestors held conferences to discuss the advisability of such and such arrangements of lever and fulcrum in lifting a weight. Of course such possible discussions never were and never could be settled until the crucial measurement—the *direct* measurement was made. It would be of inestimable value to know whether the presence of certain books in the schoolroom, or the possession of a certain amount of professional training on the part of the teacher and the like are prerequisites of certain defined changes in pupils. Without such *ancillary* measurements by teachers and supervisors, the conditions for pupils' growth cannot be arranged in advance with certainty. But we shall not arrive at such knowledge except through direct measurement. We certainly cannot claim to know the exact causal connection between defined changes in pupils, and most of the paraphernalia with which the pupil is now surrounded. In spite of our ignorance of these causal connections, the chief method of supervision at present is to attempt to judge the presence or absence or amount of presence of these levers and fulcra.

THESIS 16. MEASUREMENT IS NO RECENT EDUCATIONAL FAD

Judging from the vituperation that has been heaped upon it, and the efforts that have been necessary to propagate it, one would think that scientific measurement was something absolutely novel. As a matter of fact, educators are, and have always been confirmed users of measurement—measurement of a kind. For several generations teachers have been employing tests which, to the uninitiated observer, would differ from standard tests in only one respect. The teacher's test is usually written on the blackboard while the standard test is usually printed on paper. Present a standard test to a teacher or principal who never heard of one, and neither will recognize that it is possessed of any peculiar virtues or patent dangers. Ayres tells us: "If Dr. Rice is to be called the inventor of educational

measurement, Professor E. L. Thorndike should be called the father of the movement." And yet, if the great majority of us had thought of standard tests before Dr. Rice, or scaled tests before Dr. Thorndike, we probably should not have deemed the ideas to be dangerous.

The writer's experience with the critics of standard tests convinces him that these critics have but two important objections, first, tests are not available for measuring all the aims of instruction, and, second, tests are sometimes misused. The first objection calls for, not the disuse of tests, but greater zeal in the extension of tests. The second objection calls for zeal, not against tests, but against their misuse. The closest students of scientific measurement are rarely its opponents and at the same time they are its severest critics. They are the severest critics because their criticisms are pertinent and because they are aware of numerous defects invisible to the casual observer.

Measurement in education did not suddenly leap into existence. It has had a gradual evolution, or rather it has been on a plateau for centuries. A student's theme informs us that: "Educational measurement is ancient as a fact, medieval as a process, and modern as a science. Half of Solomon's proverbs are tests for wisdom." The Chinese had a far-flung system of testing which was a sort of beginning for the Hillegas *Scale for Measuring English Composition*. The Roman father considered his son's literary education finished when his son could read the Roman Law from the tablet in the public forum. Little progress was made beyond the conventional, formal examination until 1894. Rice conceived the idea of a comparative test to be used in measuring the results of instruction in many schools. Out of the comparative test grew norms, for the use of a comparative test upon many schools yields norms. It was the genius of Thorndike that made possible the next advance. Utilizing the Cattell-Fullerton equal-distance theorem, he devised a scale unit for the measurement of educational achievement. This marks the beginning of scientific educational measurement. Stone's *Reasoning Tests in Arithmetic*, worked out under the direction of Thorndike and published in 1908, represent a sort of transition from the Rice comparative tests to the Thorndike *Handwriting Scale* published in 1909. Subsequent students of

Thorndike's have elaborated the statistical technique for the construction of educational scales. Hillegas, Buckingham, Traubue, and Woody constructed respectively the Composition Scale, Spelling Scale, Language Scale, and Fundamentals of Arithmetic Scale.

The movement for the scientific measurement of education has spread with great rapidity. Curtis has been particularly successful in disseminating an interest in tests. Hence it is appropriate that he should have directed the testing in the first formal survey where tests were employed. The survey was the New York City Survey of 1911-12 and the tests used were the *Curtis Arithmetic Tests*.

But measurement continued to be a matter for experts because scale scores were difficult to compute, and were generally incomprehensible.

To overcome this difficulty the writer developed and popularized a plan for having all tests yield comparable and easily understood age scores such as reading age, arithmetic age, educational age, mental age, promotion age, and the quotients such as reading quotient, educational quotient, and intelligence quotient. This made measurement popular with teachers.

Later the writer invented the grade scale yielding G scores. These proved to be so popular that they came into almost immediate use on most tests from New York to Nanking. Objectively-scorable tests yielding age scores or G scores (to be explained later) gave measurement to the millions and provided the large profits which permitted the early test publishers, namely, Bureau of Publications at Teachers College, Public School Publishing Company, and the World Book Company to expand rapidly their test publication programs.

THESIS 17. TEACHERS SHOULD COÖPERATE IN ALL TESTING AND SHOULD BE ALLOWED TO ADMINISTER AND SCORE INTELLIGENCE AND EDUCATIONAL TESTS AND INTER- PRET RESULTS

Many years ago certain specialists sought to secure a monopoly of the privilege of using standard tests by trying to persuade educators to regard the tests as possessing certain mystic properties. A few of us with Promethean tendencies set about

taking these sacred cows away from the gods and giving them to mortals. Can teachers be trusted with tests? If not, then teachers ought not to be trusted with ninety per cent of their present functions. We now entrust them with the far more difficult task of teaching reading, creating concepts, and building ideals. Let us not strain at a gnat when we have swallowed fifty camels, several elephants, and a brontosaurus or two!

BOOK TWO



CRITERIA FOR THE SELECTION AND
CONSTRUCTION OF STANDARD AND
TEACHER-MADE TESTS

CHAPTER II

HOW TO SELECT AND CONSTRUCT TESTS— VALIDITY

There are in the United States about 700,000 elementary school, high school, and college teachers. It is a conservative estimate that each teacher gives on the average twenty examinations a year. This makes 14,000,000 examinations each year. The time required to construct, give, and score each examination will average, say, three hours. This means that annually about 42,000,000 hours are spent examining pupils. Even though our estimate is doubly generous, the hours would still be sufficient to show the enormous importance of examinations. Without a doubt, examinations are and will be for some time and may possibly always remain the most important form of educational measurement.

Since teacher-made tests are of this great importance educators should apply to them the same criteria that careful specialists in standard tests employ. To facilitate such application both standard tests and teacher-made tests are treated as though they were identical, which, in essentials, they are. Occasional criteria are applicable to standard tests only, but these will be so obvious that it will seldom be necessary to point them out to the reader.

Tests Should Be in Harmony with the Philosophy of Education.—Examinations, like teaching, should be an outgrowth of the educational philosophy of the school. Examinations powerfully determine both the amount and direction of pupils' effort. The amount of effort is affected somewhat by the nature of the examination. The direction of the effort, a more vital matter, is seriously conditioned. If all examinations are deemed incompatible with the accepted philosophy of education, examinations should be abolished.

Many subsequent criteria assume a certain, generally held, though not always practiced, philosophy of education.

Tests Should Lead toward Improvements in Curriculum.—The attempt to select, prepare, and apply an *adequate* program

of testing is probably the best single approach to a critical appraisal of a school's curriculum. The curriculum should lead into measurements and the process and products of measurement should in turn probe the whole educational program. In short, the tests should consist entirely of items that will influence the pupil, teacher, supervisor, administrator, and the curriculum toward good education.

Test Selection or Construction Should Be Preceded by Careful Curriculum Analysis to Discover and Clearly and Definitely Formulate a Full List of Objectives.—Objectives should be formulated clearly and definitely in advance of test construction. Given such a list of objectives, the program of measurement should be checked against it to determine its adequacy.

Tests Should Measure Organization of Memoriter Learnings.—It is doubtful whether the testing of mere memoriter learning of any sort is educationally defensible; Terry¹ has shown that the manner of the test influences the method of study, and properly warns that the teacher should be as much concerned with how students study as with what they study. Mere memoriter tests encourage mere memoriter learning.

Perhaps the simplest way to construct tests which encourage excellent methods of study is to make it a rule to ask no straight information questions at all, but rather to ask questions which require some application or integration of information. Thus, in geography, the teacher might sketch on the blackboard a continent that never was, labeling main areas by letters, and then ask such objective questions about it, as: Through what port does most commerce flow? Is the largest city at point A, B, C, or D? Is area F, M, P, or R most heavily forested?

Here the facts and principles are not provided but must be drawn from the student's reservoir of knowledge. Sometimes, however, the facts and principles are supplied by the examiner. Thus a student may be asked to indicate which of three historical events was most important and, in addition, indicate which of three reasons explains its importance.

A straight knowledge test is justifiable probably only when it

¹ Terry, Paul W., "How Students Study for Objective and Essay Tests," *Elementary School Journal*, April, 1933, and "How Students Study for Three Types of Objective Tests," *Journal of Educational Research*, January, 1934.

can be shown that such a test is an excellent index of more important integrated abilities, and when knowledge itself may be assumed to be equivalent to action of marked social significance,—as for example, knowledge of how to stem the flow of blood.

Also it may be demonstrated some day that an information test based on a random sampling from a complete dictionary or encyclopedia (but not school textbooks) yields a fair index of the excellence of the educative process and the degree of a pupil's total attainment.

Tests Should Tap Varied Types of Thinking.—When constructing thought questions whether of the multiple-choice or essay type it would be well to keep in mind the following twenty different types of thought questions listed and defined by Monroe and Carter, following their study of this matter:

1. Selective recall.
2. Evaluating recall.
3. Comparison of two things—on a single designated basis.
4. Comparison of two things—in general.
5. Decision.
6. Cause or effect.
7. Explanation.
8. Summary.
9. Analysis.
10. Statement of relationships.
11. Illustrations or examples.
12. Classification.
13. Application.
14. Discussion.
15. Statement of aim.
16. Criticism.
17. Outline.
18. Reorganization of facts.
19. Formulation of new questions.
20. New methods of procedure.

Tests Should Measure the Degree to Which Each Pupil Has Attained All the Objectives of Instruction.—There is a pronounced tendency for examinations to stress, for example, the acquisition of scientific method itself, or the possession of relatively inert knowledge of civics but not the possession of desirable civic attitudes.

It is, of course, more difficult to measure mastery of scientific

method and possession of civic attitudes, but it is better to measure them with low accuracy than not to measure them at all. The proposition might perhaps be defended that it is even better, because of the influence on a pupil's values and incidentally the teacher's, to measure such traits with no accuracy than not to measure them.

Objectives for which tests and examinations should be developed are, to use social studies as an example: (1) use of library, (2) use of reference and other books, (3) finding information and knowledge of dependable sources of information, (4) reading maps, cartoons, tables, diagrams, and charts, (5) outlining, (6) summarizing, (7) attitudes, unless they are controversial, and then acquaintance with controversial attitudes and the bases of each, (8) discrimination between what is known and what isn't, (9) strength and direction of interests, (10) ability to generalize from data, (11) ability to make application of generalizations to novel problems, (12) ability to make deductions from generalizations, (13) ability to design valid experiments, (14) ability to discover errors in thinking, (15) ability to make wise decisions in social situations, and so on.

Tests Should Indicate the Extent to Which Each Pupil Has Attained Each Objective to the Degree Proper for Him.—Book Four shows how to fix objectives or expectations in terms of the pupil's intelligence and environment. But adequate guidance of the pupil requires even more. In addition to his intelligence and background, his special aptitudes and his constantly shifting interests and purposes should be taken into account.

Tests Should Yield Diagnostic Data.—Every power test or difficulty test though not designed primarily for diagnostic purposes, yields some diagnostic information. Other things being equal, the more information of this character that is yielded the better the test.

But tests may be prepared or selected which give little attention to the increase in difficulty of the different items. Diagnostic tests aim rather to discover whether pupils have mastered all phases of a given skill or particular aspects of some subject. In such tests, a pupil's performance on a particular item or group of items is of more consequence than his total score.

Since diagnostic testing should usually be a continuous proc-

ess and coterminous with teaching itself, diagnostic tests are treated more fully in later chapters in conjunction with diagnostic teaching and test lessons.

Tests Should Be Enjoyable to Both Pupils and Teacher.—Pupils prefer new-type tests to essay examinations. It is agonizing for a pupil to describe at great length a knowledge which he does not possess in hopes that his command of English will camouflage his lack of information. Here is a question which was asked in a recent examination in educational measurement. "Which three of the tests described by Whipple do you think would be of most service in an elementary school, if your school had a school psychologist to apply them?" Consider the perspiration it must have cost a student to perpetrate this answer:

The tests described by Whipple embraced most of the difficulties that would be embraced in problems of classroom instruction. I think his tests embrace a great variety of methods of approach and it seems difficult for me to think of just three to whom the presence of a psychologist in a school would give help. I would think it would be the tests in which knowledge of the workings of a child's mind and its growth and development would be most apparent since those not particularly trained might focus on others not of this kind. I fear it would be unwise to specifically mention just three when the number is so great which would fulfill all these requirements. Every teacher to be a psychologist would help all classroom measurement work of whatever kind greatly, I know; since we cannot know of the influence of a test upon any group except by the mental reaction produced.

The multiple-choice examination is more enjoyed by the teacher. The scoring is easy, rapid, and automatic when she does the scoring, and far more rapid when the pupils do the scoring. The pupils cannot well assist in scoring the traditional examination, and for the teacher to score forty verbose examination papers is time-consuming drudgery. Every moment of the time while scoring, the teacher must be profoundly concentrating upon what she is reading, for much of the time she must be separating the chaff from the wheat where the chaff is cleverly painted to look like wheat. And along with this is a continual emotional strain caused by her resistance to the temptation to underscore some and overscore others.

An investigation by Somers¹ has made it all too clear that

¹ Somers, Grover T., *Students Attitude toward Examinations*, Bureau of Cooperative Research, Indiana University School of Education, Bloomington, Indiana.

although pupils prefer new-type objective tests and consider them better tests, pupils do not like any kind of examination very much. Examinations are regarded by them as unacceptable, something forced upon them, and on a par in satisfyingness with menial labor. According to Somers, they are prone to regard examinations as ends in themselves, as serving mainly to provide term marks, as having too much importance attached to them, as being scored inaccurately and with prejudice for pleasing personalities and verbosity, and as testing too much the ability to spot the instructor.

It may be asking too much to suggest that examinations be made enjoyable. But even those who hold that the school has a job to do and must do it, whether the pupils like it or not, will generally agree that all possible steps should be taken to make examinations as acceptable as possible. These questions are worthy of intensive study by educators: How can examinations be made more acceptable to pupils? Is there any reasonable prospect that we can make them acceptable without sacrificing central purposes which they now serve? If they cannot be made acceptable, should they be abolished?

But in considering this whole problem we should bear in mind the notification given to the graduating class at St. Lawrence College by Owen D. Young, that the world tests continuously and often when the testee is quite unaware of it, that this testing is so prolonged that the candidate cannot profit by lucky questions, and that the student should test himself often and searchingly, for though he may be justified in fooling others he cannot afford to fool himself.

Tests Should Contain Only Items of High Validity.—An item is valid if it measures what it purports to measure. Of two items which *appear* to be equally valid, one may discriminate between a “good” group, i.e., pupils who possess much of the trait in question and a “poor” group, i.e., pupils who possess less of it, whereas the other may not discriminate at all or even discriminate negatively. Furthermore, an item may discriminate for one age or grade group but not for another.

As an illustration of how an item’s validity may vary from group to group, the reader is asked to cross out the one word in these five words which does not belong in the group: *needle scissors paper thread cloth*. The reader, if he is intelligent (!)

probably crossed out *paper*, but intelligent pupils in the elementary schools tend to cross out *cloth*. If *cloth* were scored as correct for adults, the item would have a negative validity for them. Children associate *scissors* with *paper*, whereas adults associate *scissors* with *cloth*.

If the permanence of the test justifies the labor, a validity index may be computed for each item for small groups of items. McCall, Long, Vincent, Pearson, and others have invented item-validity formulae for single items, and Barthelmess in a Ph.D. thesis entitled *The Validity of Intelligence Test Items* evaluated several of them. Thorndike in his *Measurement of Intelligence* computed instead the validity of several items taken together.

Unfortunately the excellence of a test as a whole is not indicated exactly by the average of the indices of discrimination for all the items, since the inter-relationship among the items is a factor in the total worth of the test. But since the labor of computing these inter-correlations is prohibitive, we are forced to depend solely upon each item's correlation with a criterion ignoring the pattern it makes with other items, except for dubious logical checks to insure that items fairly sample the total trait and don't overlap too much.

In selecting tests the task is simpler, since here all that needs to be known is the correlation of the total scores on the test for a proper group of pupils with criterion scores for these same pupils.

Test Items Should Be Tested by Interviews with Pupils and Others.—The interviewing of enough pupils to discover *all* the mental processes that are evoked by a particular item is very, very important. It is quite impossible for the best test maker in the world to foresee all the irrelevancies which an item may evoke. Follow the best test builder and you will find him spending much time testing pupils—low, average, and superior—and then asking them one by one why they answered each item as they did. If something in the item impels low-ability pupils toward the right answer and high-ability pupils toward the wrong answer, the item is revised and tried on another group.

Sometimes, for example, he finds that in an effort to produce a difficult item, he has made the wrong choice too plausible for superior students, thus drawing them away from the right re-

sponse because they have a little knowledge only, whereas less able pupils, having none of this knowledge, are untempted by it.

Sometimes he finds that no pupil has sufficient knowledge to choose the correct scholarly response but does have a popular knowledge or prejudice which leads him toward a wrong choice, whereas the less able pupil, innocent of even the popular knowledge, is thereby given an advantage.

Sometimes he finds that the item measures intelligence when he wants it to measure some phase of achievement. To check on this point, he tries his tests or examinations on intelligent persons who know nothing of the subject. Items on which they do well are discarded, or revised if removable clues capitalizable by intelligence are discovered. Such clues may be grammatical consonance between the preamble and the choices, conventional modes of expression, consonance with the maker's general, known point of view, and many others.

Test Items Do Not Have to Appear Valid to Casual Inspection or Even Expert Inspection.—It should not be necessary for test makers to be compelled to discard excellent items because of their external appearance. The items in the *Comprehensive Achievement Test*, for example, discussed in Book Four were prepared with conscientious care. Every item was carefully formulated, criticized by two specialists, revised, criticized by teachers and supervisors with varying philosophic viewpoints, revised, tried on pupils who were interviewed, revised, and tried again. Sometimes the items which came through these trials and were adopted for final use possessed certain external appearances which the authors knew would be an open invitation to criticism by untrained persons who think a good item can be told from a bad one merely by looking at it. Other good items were regretfully discarded because, if included, the authors could not go along with the test and explain the intricate considerations which led to their inclusion. Every test user should have the experience of watching such a test being constructed and see how often items, seemingly good at the start, are altered from their initial form. They would then realize how deceptive are appearances, how earnestly the specialist strives to get items, which will function no matter how they appear, and how important it is that they be loth to criticize

until they have collected better data than the specialist was able to secure and have tried to make better items.

Tests Should Use Whatever Testing Technique Is Most Relevant to the Trait and Group Being Measured and the Purpose the Measurement Is to Serve.—Many testing techniques have been invented of which only the most common need to be mentioned here. Illustrations of all types discussed, except varieties surely familiar to the reader, may be seen in Chapters V, XIV, and XVIII.

The most common test types are: the *True-False* or *Yes-No* (discussed in detail under the next criterion), the *multiple-choice* (an item with one right choice to be indicated mixed with two wrong answers, one right and three wrong answers, or one right and any number of wrong answers, although in practice five wrong answers is about the limit), the *multiple-response* (an item with two, three, or four right answers, to be indicated mixed with one, two, three, or more wrong answers), the *matching* (an item containing two or more sub-items in one column each of which is to be properly matched with some sub-item or sub-items in another column), the *completion* (a statement from which words have been deleted and are to be replaced by the pupil, the words to be chosen from a list provided or produced mentally by the pupil), the *simple-recall* or *one-word answer* (an item which requires the pupil to recall a name, date, or the like, or provide the answer to a problem in arithmetic), and the *essay*—an item familiar to all.

The correct and incorrect choices in either the multiple-choice or multiple-response types may each complete a preceding sentence, or may appear in the middle of a sentence, though the last is not advised for general use.

Of all these, the completion test is the least useful as an achievement test, because it tends to encourage purely verbal learnings; often suffers from having too many words removed, thus making it primarily an intelligence test; is easily mutilated at the wrong place; and is difficult to score either objectively or conveniently unless the list of completions is provided.

Due to the extreme subjectivity of scoring it the essay examination should be used only when it is desired to test a pupil's ability creatively to organize or summarize a complicated subject and word it effectively.

As Lindquist ¹ has so well said in an excellent treatment of this subject, there have been too many tests of the *who*, *what*, *when*, *where*, and *describe*, *define*, and *name* varieties and not enough of the *how*, *why*, *wherefore*, *with what consequences*, of *what significance* or the *explain* or *interpret* types. This has come about through an effort to make it easier to score essay examinations. The author agrees with Lindquist that it is far preferable to secure ease of scoring by the use of the multiple-choice type of test which readily lends itself to the interpretative and applicational aspects of a subject, especially if the choices offered are a paragraph or more in length.

The matching type of test is seldom used because it cannot conveniently be employed on as many kinds of subject matter as can the multiple-choice form and because it is more irksome to pupils. This is true whether the two sets of items to be matched are equal or unequal (latter is preferable), or whether some items are to be used more than once in matching (a good plan), or whether the items are homogeneous in character (to be preferred). If used, this type of test can be made less irksome and time-consuming by placing the short set of items (if one is shorter) in the right column, and by sequential arrangement of items if they are alphabetical, chronological, or the like.

The multiple-choice type, whether two-choice (true-false or yes-no), three-choice, or four-or-more choice is the most useful and popular of all types. If this type is used it is desirable that all choices be good enough to require that all be read, that the correct answer should not be distinguishable because of length or other superficial factors, and that the items be so phrased that the choices appear at the beginning or, better, the end.

The simple-recall type should be preferred to the multiple-choice type when it is important (which it rarely is) to test recall rather than recognition, when there is just one possible brief answer, as in the case of arithmetic problems, when absolutely objective scoring is not imperative, and when electric machine scoring is not planned. Since there is little or no guessing by the pupils, fewer items are required to make the test reliable.

Again, certain traits are more appropriately tested by such testing techniques as oral questioning, interview, library assign-

¹ Hawkes, Lindquist, Mann, and Others, *The Construction and Use of Achievement Examinations*, Houghton Mifflin Co., Boston, 1936.

ment, rating, observation of behavior, attitude scale, and interest questionnaire.

Then, too, the group to be measured influences the choice of technique. At present some believe that attitude scales and interest questionnaires cannot be used readily with pupils below the junior high school level.

Finally, consideration must be given to what purpose or purposes the test will serve, whether for grading, marking, motivation, diagnosis, guidance, *et cetera*. And purpose affects not only choice of technique but every other aspect of measurement.

Test Items Should Be Brief to Aid Adequacy of Sampling.—Of all the types of tests, the True-False or Yes-No variety permits the widest sampling in a given time. The scattered examination shown below is designed to test a pupil's knowledge of certain facts concerning the physical features of the United States and to do it by means of an objectively scorable, brief-item test. In actual practice a teacher will usually test on a much narrower topic. We have purposely written this examination hastily in order that it might illustrate certain crudities of construction. Any teacher in the elementary school could do as well and most teachers could do better. The same technique is equally useful to high school and college teachers.

The examination as presented here assumes that the statements whose truth and falsity are to be determined by the pupils have been mimeographed so that a copy of the examination could be placed in the hands of each pupil. It could instead be written on the blackboard or dictated orally. The sample examination given below has been worked through by a pupil and been scored by a pupil or the teacher. The underlining was done by a pupil. The check, cross, and zero mean respectively that the pupil's answer is correct, incorrect, or omitted. Only enough of the examination is shown below to illustrate the procedure.

SAMPLE EXAMINATION ON THE GEOGRAPHY OF THE UNITED STATES

Some of the following twenty statements are true and some are false. When the statement is true draw a line under True; when it is false draw a line under False. Be sure to make a mark for every statement. If you do not know, guess.

1. In general the mountain ranges run east and west.	True	<u>False</u>	✓
2. Most of the rivers flow north.	True	<u>False</u>	✓
3. Mt. Mitchel is the highest point east of the Mississippi River.	True	<u>False</u>	X
4. Mt. Washington is higher than Mt. Mitchel.	<u>True</u>	<u>False</u>	X
5. The Catskill Mountains are in Maine.	<u>True</u>	<u>False</u>	✓
6. The Cascade Mountains are nearer the Pacific Ocean than the Rocky Mountains.	True	<u>False</u>	X
7. The Rocky Mountains are nearer the Pacific Ocean than the Appalachian Mountains.	<u>True</u>	False	✓
8. The Blue Ridge is in the Rocky Mountains.	True	<u>False</u>	✓
9. There are more active volcanoes in the west than in the east.	<u>True</u>	False	✓
10. "Old Faithful" is the name of a cyclone which sweeps upward from Texas into Oklahoma.	<u>True</u>	False	X
11. The "Grand Canyon" was cut through the Cumberland Plateau by the Susquehanna River.	True	<u>False</u>	✓
12. Pike's Peak is in the Rocky Mountains.	<u>True</u>	<u>False</u>	✓
13. The Mississippi River flows into the Great Lakes.	True	<u>False</u>	✓
14. All the following are tributaries of the Mississippi River: Arkansas, Missouri, Ohio.	<u>True</u>	False	✓
15. The Big Sandy is the biggest river in the United States.	<u>True</u>	False	X
16. The Atlantic Ocean is to the east and the Pacific Ocean to the west.	True	False	O
17. Canada is to the south and the Gulf of Mexico to the north.	True	<u>False</u>	✓
18. The great lakes are five in number.	<u>True</u>	<u>False</u>	✓
99. It is easier to sink while swimming in the largest lake east than in the largest west of the Mississippi.	<u>True</u>	False	✓
100. The central portion of the United States is on the whole more level than the eastern or western portion.	<u>True</u>	False	✓

It is claimed that this type of examination does not require the pupil to demonstrate a power to organize his materials. This is true in the sense that the pupil does not *describe in writing* a complicated mental organization but a statement *can*

be so worded as to require an exceedingly complex mental organization before a correct answer can be unfailingly given. Consider the mental organization that must precede a correct answer to this simple statement: "If the trade winds blew east Peru would have luxuriant flora." If it is desired to test a pupil's power to word his thought a composition test may be given.

Again, it is claimed that this type of examination can test knowledge but not skill, knowledge but not the ability to do. Even skills can be tested by this examination. To reason that trade winds blowing east would be warm, would absorb moisture from the Pacific, would become chilled in passing over the Andes, would consequently deposit a heavy rainfall for Peru, which taken in conjunction with the equatorial climate would produce a luxuriant flora, is one sort of skill which this examination will test. Mathematical skills and the like may be tested in at least two ways, though there are better ways. An example or problem may be stated together with an answer. The pupil's task would be to determine by working the problem whether the answer given is true or false. Or instead, the teacher can work the problem on the blackboard for all the pupils and have them indicate whether her process was correct or incorrect.

Because the True-False test may be made more representative of the total field of the pupil's study, it is a fairer measure of the pupil. In the case of the traditional examination the teacher is forced to select a very small number of questions. When we were students almost as much of our ingenuity went into divining the kind of examination questions the teacher would ask as in reviewing for examination. Now that we are teachers we have no reason to suppose that this practice has ceased.

Tests Should Provide a Comprehensive Measure of the Trait.—Comprehensiveness is feasible when the examiner is interested in only a narrow ability or limited field of subject matter. Some more economical method must be found for measuring a comprehensive ability.

A test can be made comprehensive by including random samplings of the ability in question. In order to determine how many words a pupil can spell, or define, or use, it is not necessary to try him on every word in Webster's Dictionary. It can be done just as well by taking from the dictionary a random sampling of its words. In making such a sampling it is impor-

tant that the samplings be made random, and that enough samples be employed to yield a reliable measure of the pupil. Randomness may be secured by using the first or ninth or any other numbered word on each page or each third page or each twenty-fifth page or the like of the dictionary. This will suggest how chance samplings may be made from a variety of subject matter. It is worth pointing out that when test material is selected according to this random-sampling method, the construction of duplicate tests becomes a very simple matter. The value of such duplicate tests will appear later. It should be remembered that the method of random sampling answers only the question: What per cent of a total field of knowledge does a pupil know? Except for the elements in the test, such a test leaves us in ignorance as to just what elements in the field of knowledge the pupil knows.

To overcome this last obstacle, especially in the field of skill tests, it has been suggested that comprehensiveness be secured by using type material. This type principle of selection assumes that each subject involving skill contains typical units or typical processes, and that the pupil's ability in the entire subject is substantially determined by measuring his ability in the type processes. The fundamentals of arithmetic, for example, are supposed to contain certain type processes. The ability to *carry* in addition is one such type process. The ability to fix the decimal point in division is another type process and so on. It is held that a test to be representative of the fundamentals of arithmetic must contain every type process.

Tests Should Subordinate Statistical Considerations to Diagnostic and Social Significance.—Monroe¹ has criticized the *Woody Arithmetic Scales* because Woody did not select examples for his tests primarily on a type basis. Monroe contends that Woody sacrificed diagnostic ability to statistical beauty, since Woody retained examples in his scales primarily because of their statistical behavior—because of their difficulty.

Another principle for selecting test material which has come into common use is the social-worth principle. This principle makes comprehensiveness or difficulty subordinate to relative value. The social-worth principle assumes that the most valu-

¹ Monroe, Walter S., "An Experimental and Analytical Study of Woody's Arithmetic Scales"; *School and Society*, October 6, 1917.

able information for the school will come from testing the pupil's ability to spell only those words, or solve only those problems, or demonstrate a knowledge of only those historical facts which are of greatest social value. The best illustration of a test whose construction has been guided by this principle is the *Ayres Spelling Scale*. The Ayres test contains 1000 words which were selected by exhaustive investigations to discover which words were most frequently used. Similar surveys for other subjects have made it possible to construct other tests in accordance with this principle.

Comprehensiveness requires that we not only measure how much a pupil can do and how well he can do it, but also we must measure how rapidly he can do it. This proposition needs no justification, for the practical importance of such a diagnosis of the pupil's habit of work is obvious. At least one major aim of the school is to prepare the pupil for effective participation in the social group. The social group does not want the pupil's ability, nor does the pupil derive much joy or profit from his ability, if he falls below a minimum of speed. Thus the three main dimensions of a pupil's ability are (1) how much or how difficult, (2) how well or how accurately or with what quality, and (3) how rapidly. If reading is to be measured, a test or tests (for frequently all three dimensions cannot well be measured in a single test) should be selected which will measure all three aspects of reading.

Tests Should Be Free from Irrelevancies.—Test results are more comparable to life results when they are free from irrelevancies. To return to the illustration of a reasoning test in arithmetic, the arithmetic problems probably more nearly duplicate real problems when they are free from non-arithmetical difficulties. Complicated instructions for the test might so confuse the pupils as to leave no fair opportunity to attack the arithmetical difficulties. Again, a complex wording of the problems might make the linguistic difficulty of greater consequence than the difficulty of the mathematical processes themselves. In selecting tests they should be carefully studied to discover whether everything possible has been done toward the elimination of irrelevancies in instructions and in the organization and wording of the test elements, or at least toward determining the influence of these irrelevancies.

While linguistic irrelevancies are more common, they are not the only kind by any means. The form of the test is often an irrelevancy. Not only must the pupil overcome the difficulties of the real test material, which is always to some extent camouflaged by linguistic irrelevancies, but he must also overcome the difficulty of the general form in which the test is couched. These moulds for test material are many. There are the *question* mould, *completion* mould, *classification* mould, *matching* mould, and many others. All these irrelevancies are important elements of difficulty especially for young children. They do greatest harm in rate tests where the speed score of the pupil is much influenced by the rapidity with which he adapts himself to the test.

Terman¹ says of the army intelligence test Alpha: "The test questions were ingeniously arranged so that practically all could be answered without writing, by merely drawing a line, crossing out or checking." There were various reasons for this provision, such as to require less time for testing and to make scoring economical and objective. But a very important reason was to make a test which would test the thing for which the test was designed. It was designed to measure general intelligence. If writing were made a prominent feature of the test, the test would tend to give a measure of speed of handwriting rather than of intellectual ability. Individuals are more alike in their speed of checking, crossing out, and underlining than they are in speed of penmanship.

It is possible, especially in the case of very long tests, that the chief factor measured is not the ability desired but fatigability. The test should be of such a length or so constructed as to eliminate fatigue, particularly if some of the pupils fatigue more easily than others. This point needs most attention when comparisons are to be made between young and old children.

Fatigue may be eliminated in various ways. First, the test may be made short. Second, if reliability requires a longer test, the test may be divided into parts with a rest or exercise interval between. Third, if the test consists of a series of short tests, the shorter tests may be so arranged as to have difficult tests followed by easy tests and tests of one nature followed by tests of another nature and *vice versa*. Fourth,

¹ *Psychological Bulletin*, June, 1918.

the test may be made variegated and interesting both as to type and material. The material in the Alpha intelligence test for the army, for example, kept the recruits in a merry and at times almost boisterous mood throughout.

The foregoing propositions concerning irrelevancies should be accepted with caution and applied with care. The propositions were made more to direct attention to certain problems rather than because they have a firm experimental basis. If the examiner's purpose is to make a psychological study of pure arithmetical abilities there can be no question but that every possible linguistic or other irrelevancy should be eliminated from the tests used. Similarly when linguistic ability is being measured, all non-linguistic difficulties should be eliminated. But if life's arithmetic problems are to be duplicated we cannot be so sure of the value of eliminating all irrelevant difficulties. When a child pays for purchases in a store he must steer his course through numerous distractions which are not all mathematical in their nature. Since these practical distractions cannot conveniently be duplicated in a test, perhaps the linguistic or other difficulties should be retained as a sort of substitute. Again, the propositions should be applied with care, because an irrelevancy in one test may not be so at all in another test. If the form or mould of a test duplicates the pattern of the pupil's mental processes in performing an actual task, the form of the test is not an irrelevancy. A casual inspection of the following task taken from the *Woodworth-Wells Directions Test* would give one the impression that the whole test is nothing but an irrelevancy, and yet this impression would be a mistake, for the purpose of the test is to measure the ability to deal with just such complicated directions.

With your pencil make a dot over any one of these letters **F G H I J**, and a comma after the longest of these three words: boy mother girl Then, if Christmas comes in March, make a cross right here., but if not, pass along to the next question, and tell where the sun rises. If you believe that Edison discovered America, cross out what you just wrote, but if it was someone else, put in a number to complete this sentence: 'A horse has feet.' Write *yes*, no matter whether China is in Africa or not.; and then give a wrong answer to this question: 'How many days are there in a week?' Write any letter except *g* just after this comma, and then write *no* if 2 times 5 are 10 Now, if

Tuesday comes after Monday, make two crosses here ;
 but if not, make a circle here or else a square here
 Be sure to make three crosses between these two names of boys: George
 Henry. Notice these two numbers: 3. 5. If iron is
 heavier than water, write the larger number here , but
 if iron is lighter write the smaller number here Show by a
 cross when the nights are longer: in summer? in winter?
 Give the correct answer to this question: 'Does water run
 uphill?' and repeat your answer here Do
 nothing here ($5 + 7 =$), unless you skipped the preceding
 question; but write the first letter of your first name and the last letter
 of your last name at the end of this line:

Tests Should Exclude Ambiguous and Negative Items.—Statement number 18 in the sample True-False test is somewhat ambiguous. It says: "The great lakes are five in number." Since *great lakes* is not capitalized a pupil might very legitimately interpret this to include the Great Salt Lake and others. It will later be difficult to satisfy this pupil that his score should suffer because of the construction he gave this sentence. If the teacher will study her mistakes in this respect she will soon learn how to reduce such ambiguities. As any teacher can testify, the danger of ambiguities of wording are not peculiar to this test. This type of test does not, however, give a pupil an opportunity to reveal just what interpretation he places upon each statement unless the teacher follows the procedure of having pupils score their own or each other's paper. Self-scoring will reveal all cases of ambiguity. Statements which are particularly flagrant in this respect can be omitted in scoring.

Statement 99 in the sample test illustrates another irrelevancy. The purpose is to test whether the pupil knows that the largest lake west of the Mississippi River contains more salt than the largest lake east of the Mississippi. Instead of measuring this the item may be testing whether a pupil knows that it is easier to sink in fresh water than in salt water. Complex wording, unfamiliar terms, the use of negatives, all tend to make the test a linguistic one. Simple, brief statements without negatives are best. Brevity is particularly important if the test is to be administered by reading it aloud.

Tests Should Avoid Suggestive Items.—The teacher may so construct the examination as to force pupils to guess wrong due to the power of suggestion. This probably explains why state-

ment 15 was marked wrongly. The pupil doubtless argued to himself that since the river is named the Big Sandy it probably is the biggest river in the United States. The influence of having many suggestive statements in the test is to make the examination more difficult. It operates to give to the pupil who knows nothing at all in the test a large negative score instead of a zero score and it penalizes rather heavily the pupil who does much guessing, for every time he allows himself to be suggested in the wrong direction a point is subtracted from the score he has already made by what knowledge he has. In other words, the suggestive statements make the gap between those who know much and those who know little wider than it otherwise would be. Whether a pupil should be specially penalized for yielding to suggestion is an arguable question. There may be situations where it is eminently desirable to determine whether pupils know what they know so well as to be able to resist suggestion. In general, however, it is best to avoid suggestive statements. The ideal should be to construct the examination so that any pupil who knows absolutely nothing about the test will make a score of zero.

Tests Should Contain, If They Are True-False Tests, Approximately the Same Number of True and False Statements.—

A clever pupil may get a higher score than he deserves if he discovers there are many more true statements than false statements in the test or *vice versa*. Suppose there are many more true statements than false statements and suppose some pupil discovers this by observing the statements that he knows, or by observing the teacher's bias for writing true statements instead of false ones. Naturally when he does not know what to mark he will mark *True*, thereby securing a larger score than his ability justifies. Probably it is by just such utilization of the errors of others that the intelligent get through life so much more smoothly than the stupid. On the other hand, the teacher should not have exactly the same number of true and false statements each time, because this will invite clever pupils to count back to see how many more true statements have been marked than false statements. Sometimes there should be more true statements, sometimes more false statements, sometimes the same number of each. Any regularity of plan should be carefully avoided. An English admiral complimented the skill of German

submarine commanders by saying they were masters of irregularity. Let chance determine what shall be true and what shall be false.

Tests Should Be So Made That Students Will Be Unable to Infer the Answer from Verbal Clues.—Weidemann, Brinkmeier, and Ruch have analyzed numerous True-False items prepared by teachers and find that a bright, discerning student would materially increase his score by noting such clues. Thus an overwhelming majority of tests items containing the word *all, always, never, no, none, nothing, only, cause, or reason* are false, whereas an overwhelming majority of items containing enumerations, *may* (expressing possibility), *most, some, often, generally*, and the like are true. Test items containing these terms should be scrutinized with care to make sure that there is an equal number of true and false items.

Tests Should Not Contain Items Which Aid the Pupil to Answer Correctly Other Items.—An astute pupil can frequently profit unduly from items which overlap in content or are otherwise related unless such items are formulated with this caution in mind.

Multiple-Choice Tests Should Have No More Choices in Items Than Are Sensible and No Discoverable Plan for the Spatial Location of the Right Choice.—Silly or obviously improbable choices that no pupil will be likely to make are largely waste from a testing point of view. They are not wholly waste since they reduce the likelihood that a pupil who is guessing will get a correct answer by chance. But it is doubtful whether this is sufficient justification for their inclusion.

There is value in having one right answer and the same number of wrong choices in all test items only when it is important to know how many items a pupil knew and how many he got right by guess. Since the teacher rarely needs to know this, the same examination may contain items with two, three, four, or more choices.

The right answer in multiple choice tests should occupy a chance position. All the right answers should not appear in the first position or second or third or fourth. Neither should they be rotated according to any discoverable plan.

Tests Should Not Contain Trivial Items Lest They Induce Wrong Habits of Study.—The greatest defect of the sample True-False test is that it contains items of negligible signifi-

cance. This can be most readily seen by comparing the items in it with the items in the *Comprehensive Achievement Test* presented in Book Four.

Tests As a Whole Should Be Valid.—When a test really measures what it purports to measure and consistently measures this same something throughout the entire range of the test it is a valid test in its entirety.

Ask a cautious psychologist just what a given test measures and he will answer somewhat as follows: "It measures the ability to do so and so with the material which you see on the test sheet, when the test is applied under certain conditions." If you are dissatisfied with this conservative statement you may enquire: "Will the pupil who deals with these test difficulties with a given degree of excellence deal with these apparently same difficulties when imbedded in a real, practical life situation with an equal degree of excellence?"

No one knows very much about just how close results for the different tests are to the results in actual practice. We give a class a paper test composed of twenty reasoning problems in arithmetic. Johnny does eighteen of the twenty problems. Had he met these twenty problems at the store or the post office or the playground, would he have succeeded with these same eighteen problems and failed on the same two? Nobody knows. If he did not do the eighteen but did do sixteen, would Mary and Lucy who did fourteen and twelve test problems respectively show proportional decreases when faced with real problems or might they possibly surpass Johnny? In other words, if test results and life results do not coincide, do they even correlate, i.e., does the pupil who makes the highest test score make the highest life score and the one who makes the second highest test score make the second best life score and so on? Nobody knows. We know enough to say that there will be a rough correspondence and probably a close correspondence, for the chasm between test conditions and life conditions does not yawn as wide as some would have us believe. It is undoubtedly wider for some tests than for others.

As Far As Practicable, Tests Should Present a Real Life Situation.—Test results are more comparable to life results the more nearly the test process approaches the character of the life process. The ability of pupils to spell, for example, may be

determined by (1) searching through their letters, compositions, and the like, (2) having them write dictated sentences in which the critical words are imbedded, (3) having them write isolated words pronounced by the examiner. The composition method more nearly duplicates the life process, the dictation method next, and the column spelling last. Again, pupils' ability in grammar can be measured by making an analysis of their written or oral compositions or by giving them a specially devised grammar test. The former test would yield more natural results. It is of course one of the perversities of fate that an increase in naturalness is attended by an increase in inconvenience.

Tests vary greatly in the exactness with which they reproduce the life process. Hollingworth¹ lists four fundamental types of tests: miniature, sampling, analogy, and empirical.

He writes that in the case of the *miniature* test the "entire work, or some selected and important part of it, is reproduced on a small scale by using toy apparatus or in some such way duplicating the actual situation which the worker faces when engaged at his task. Thus McComas, in testing telephone operators, constructed a miniature switchboard and put the operators through actual calls and responses, meanwhile measuring their speed and accuracy by means of chronometric attachments."

The *sampling* test measures a candidate's ability to do an actual sample instead of a toy representation of a given occupation. A would-be stenographer is given an actual test of ability with dictation and with a typewriter. A clerical aspirant is set to finding addresses in a telephone directory or copying a table of figures. Practically all educational tests are dummy samplings of this variety. We test a pupil's reading ability by means of samples of reading material. We test his ability to solve problems in arithmetic by giving him sample problems in arithmetic to do.

The *analogy* test employs material which is neither the same as nor similar to the material of the occupation, but it is supposed to exercise those mental traits requisite for success in the occupation. To quote Hollingworth again: "Thus girls employed in sorting steel ball-bearings, and also typesetters, have

¹ Hollingworth, H. L. and L. S., *Vocational Psychology*, D. Appleton & Co.

been selected on the basis of their speed of reaction to a sound stimulus." During the World War, Stratton, Henmon, Thorndike, and others attempted to devise tests which would be diagnostic of ability for flying. At that time no empirical tests existed, and dummy tests were impractical. So those who were working on the problem first made an analysis of the mental and physical characteristics upon which success in flying would logically seem to depend, and then devised means for measuring a candidate's possession of these traits. Tests were devised to measure a candidate's sense of balance, perception of tilt, nerve-resistance to sudden sensory shock, and the like. By checking each of these tests against subsequent success of aviators, it was found that some had no significance at all, while others were slightly symptomatic. A composite score from those tests which were found valuable, selected aviators with fair accuracy. In similar manner tests were constructed to select shell inspectors, gun assemblers, etc. Pursuing this same method of analysis, Rogers has constructed tests for determining whether pupils possess mathematical capacity. Briggs has constructed similar tests for foreign language capacity.

The *empirical* tests are those which were discovered from a more or less haphazard trial-and-error search. The test selector makes no conscious attempt to select or construct a test which is a miniature or sampling or analogy. He tries out a number of tests, eliminates those which are not symptomatic and retains those which are.

Tests Should Have Known Validity Correlations.—How may we know whether a given test measures the ability which we desire measured? We know what a test measures only by its correlations. Does a pupil's score on an intelligence test coincide with the school's and world's estimate of this pupil? Does the arithmetic test indicate how well a pupil will be able to work examples or solve arithmetical problems in the store or in those realms for which the school is preparing the pupil?

Two ways of determining this correspondence are available. One method is to give a test to a group of pupils, to preserve the records, to follow up the testing with prolonged careful observation of how well these pupils do in real situations which may or may not be arranged by the investigator, to rank the

pupils in order of their ability first on the test, and second, in the real situations, and finally, by the method of correlation or inspection to determine the correspondence between these two rankings. If the agreement is close the test does measure real ability in the sense that it can rank a group of pupils in order for their possession of the ability in question. An even more careful technique is required to determine the extent to which a pupil will make the identical score in both the test and the life situation.

The second method available is to apply the test which is being validated to a group of individuals whose real ability is already known. If the test distinguishes the different degrees of known merit, we can call the test satisfactory. Ruml, Robinson, Chapman, Meine, Kruse, Wylie, Toops, and others constructed about 100 Trade Tests for the army during the war. As the following quotation from the *Psychological Bulletin*, June, 1918, will show, they employed this second method to determine whether their tests really measured the trade skills which the tests purported to measure. Few educational tests are constructed with such careful attention to what the tests really measure. The test is usually assumed to measure what it looks as though it measures.

Evaluating the test.—If a trade test is good, a known expert, when tested, is able to answer all, or nearly all, the questions correctly; a journeyman is able to answer the majority; an apprentice a smaller part, and a novice practically none. This does not mean that each question should be answered correctly by all the experts, a majority of the journeymen, some apprentices but no novices. There are few questions which show this result.

Other types of questions, however, are more common. Some show a distinct line of cleavage between the novice and the apprentice. Novices fail, but apprentices, journeymen, and experts, alike answer correctly. There are likewise questions that are answered correctly by nearly all journeymen and experts but only a few apprentices, and questions that only an expert can answer correctly. Each type of questions has its value in a good test. The main requirement is that the tendency of the curve should be upward; a question which is answered correctly by more journeymen than experts or more apprentices than journeymen is undesirable and is at once discarded. A proper balance is made of the others.

Calibrating the test.—One task still remains; namely, that of calibrating the test. It becomes necessary to determine how many points should indicate an expert, how many a journeyman, etc. Obviously

the way to do this is to note how many points were scored by the known experts and the known journeymen when they were tested. Ordinarily the expert scores higher than the journeyman and the journeyman higher than the apprentice. It frequently happens that a few journeymen score as high as the lowest of the experts and a few apprentices as high as the lowest of the journeymen. There are consequently certain overlappings between the classes. In calibrating, the object is to draw the dividing line between classes so that the overlapping shall be as small as possible.

When these dividing lines, or *critical scores* as they are usually called, are established, the test is ready for distribution to camps.

Suppose that we give a group of pupils a test in arithmetical problems, and then, without arousing the suspicion of the pupils, arrange the situation so that these same pupils will meet these same arithmetical problems in their play life on the street, and suppose that the test and the observations upon the pupils' success with the play problems are reliable measures of each of these abilities and suppose, finally, that the correlation between the test and the observations is of only average closeness, does this condemn the test as not being a measure of real ability? Assuming that proper experimental precautions have been taken, this correlation certainly tells us that the test problems are a rough but not an accurate measure of play problems. But before we condemn the test we ought to correlate the pupils' scores on play problems with their scores on those same problems when shopping for their mothers or some other practical situation. It is not known, but it is very possible that the correlation between different real-life situations is no closer than between the test and any one of these situations. In sum, it is even probable that there is no such thing as real ability, in the sense that we are discussing it, but that there are instead, many abilities differing somewhat one from another. It is hopeless to expect to find a test which will closely correlate with each of these life situations, wrapped about, as each is, with its own individuality or specifness.

It might be possible to eliminate experimentally, all of the specifness belonging to our test and each life situation, and thus demonstrate a perfect correlation between all the thus purified abilities. Such an analysis of abilities would be of considerable theoretical interest. But for the purpose of prophesying success in life and the like, we cannot deal with these

rarefied abstractions of abilities, for abilities must always function through specific situations. It would have been of no comfort to Ruml and his colleagues to know that their Trade Tests, when experimentally purified, correlated perfectly with similarly purified trade situations. They were asked to construct tests which would, with the least error, select men who could succeed in a variety of specific situations. It is no condemnation of an educational test if it shows only substantial correlation with a variety of real situations. It is a condemnation when it shows little or no correlation with real abilities or when it shows less correlation with such abilities than some other available test which is equal in all other requirements.

If, as is often the case, the test itself provides the best obtainable measure of a given trait that competent persons can suggest, then the test itself, rather than something outside the test, becomes its own criterion and hence its validity correlation may be assumed to be perfect.

The following references deal critically with the concept of validity:

Smith, B. O., *Measurement in Education*, Bureau of Publications, Teachers College, Columbia University, New York, 1937.

Monroe, Walter S., *Introduction to the Theory of Educational Measurements*, Houghton Mifflin Co., Boston, 1923.

CHAPTER III

HOW TO SELECT AND CONSTRUCT TESTS— RELIABILITY AND OBJECTIVITY

Tests Should Be Reliable.—By reliability of a test is meant the amount of agreement between results secured from two or more applications of a test to the same pupils by the same examiner. Perfect reliability obtains when an identical examiner applies two identical or exactly duplicate tests according to an identical procedure to identical pupils. This last sentence indicates in brief those attributes which are essential to high reliability in a test, and the absence of which makes for unreliability.

One source of unreliability in a test is variation in the behavior of the examiner produced by causes external to the test itself. There are a host of causes which have the power to produce large or subtle changes in the personality and behavior of the examiner, which behavior may in turn operate to raise or lower the pupil's scores. Such possible causes are an obstreperous pupil, a welcoming smile from the teacher, an indigestible lunch, etc. Chance might produce an especially favorable combination of causes at the first testing and an unfavorable combination at the second. Such a situation would tend toward differences in results and hence toward unreliability. This cause of unreliability is not an attribute of the test itself.

A second source of unreliability in a test is variation in the behavior of the examiner produced by causes inherent in the test itself. These causes may be in the instructions for the test, the method of scoring, or the statistical treatment of results. Perhaps the most important of these causes is inadequate description. Ideally the author's description should reveal exactly how the examiner is to deal with every significant situation which may arise in the process of testing, scoring, etc. When an author begins the description of how to administer his test, in this fashion: "See to it that all pupils understand what is expected of them," there is offered an opportunity for wide variation between different administrations of the test. *Instructions*

are a part of the test and should be just about as definite and uniform as the test material itself. Definite instructions to the examiner as to how to score with uniform rigor and how statistically to treat results are no less important. A study of the extent to which *Binet Test* examiners have found it necessary to carry standardization of procedure will give a good idea of the importance of this point.

A third source of unreliability in a test is the never-ceasing moment-to-moment variation in pupils themselves. Like the examiner, each pupil is at any one moment influenced by a multitude of minute forces which pulse and play like mirrored lights on moving water. An automobile horn, the lonesome howl of Jack's dog, the bleating of Mary's lamb, a sudden thought of the swimming hole, growing discomfort of strained posture, these and a thousand other large and small internal and external influences register themselves in the pupils' scores. It is rare for the registration to be equal at two test periods, and as a consequence, results from two tests differ. It is this difference which makes the test unreliable, for there is often no reason to believe that the pupils' reactions at one test period are more typical than at another.

It is not, however, always fair to judge a test's reliability by the absolute similarity between the two scores for each pupil. There are certain *constant* causes which operate to produce absolute differences in results and hence make a test's reliability appear less than its real reliability. These constant forces must be eliminated or allowed for before the real reliability can be determined. Such constant causes are improvement due to experience with the first test and due to normal growth in the measured trait. For pupils insist upon changing with increased age and increased experience. Every second leaves its ever so little deposit. Goaded by this distracting refusal of pupils to remain stationary, Ayres has suggested that chloroforming experimental pupils would be a great convenience!

How may these constant causes be eliminated? Four methods have been employed: the methods of optimum interval, duplicate test, experimental allowance, and self-correlation. The first three methods aim to reveal the absolute similarity between the two scores for each pupil; the last method only permits a relative comparison. The optimum interval method is

zero provided the influence of these factors is equal for each pupil, which is substantially the case after any reasonable interval.

Tests Should Be Sufficiently Reliable for the Purpose.—How high should the reliability of a test be? The answer is: the higher the better. If the self-correlation coefficient is zero, the test is worthless; if the coefficient is unity, the test reliability is perfect. Here are the reliability coefficients for four standard educational tests: .7, .75, .8, and .9. All uses of test results are based upon pupil scores, and a class score which is usually a mean or median of the pupil scores. An average score for a class of ordinary size will be sufficiently reliable for most purposes even though the test's self-correlation is .7. The larger the group for which the average is computed the less the self-correlation needs to be. But if the test scores are to be used to make important judgments concerning individual pupils, the self-correlation should be above .9. Scores for individual pupils have some value, however, even when the self-correlation falls considerably below .9. A test whose self-correlation is anywhere above zero is better than nothing at all for measuring individual pupils.

Tests Should Have Enough Forms to Permit Increasing Reliability by Averaging Scores on Several Forms, Measuring Growth, and Using the Same Test Year After Year.—How may a test's reliability be increased if it falls below what is required for the purposes of the investigation? Suggestions have already been made as to how to decrease variation in the examiner and hence increase reliability through a better standardization of test procedure. An additional source of unreliability is the variation in pupils due to the operation of chance causes other than those contributed by the examiner.

There are three ways in which these chance causes of unreliability may be overcome: first, by increasing the length of the test; second, by averaging results from repetitions of the test or the test and its duplicates, and third, by a combination of the first and second methods. Unfortunately there is a limit to the number of times an identical test may be repeated owing to its increasing familiarity to the pupil, and this limit varies for different kinds of tests. In case a high reliability is desired, the existence of duplicate tests may therefore become an important

factor in determining a test's worth. Duplicate tests are equally useful in preventing coaching and in measuring growth.

Tests Should Be Objective in Administration, *et Cetera*.—

A test is perfectly reliable when identical results are secured from two applications of a test to the same pupils by the *same* examiner. A test is perfectly objective when identical results are secured from two applications of the same test to the same pupils by *different* examiners. A test is perfectly subjective when no two examiners agree. Ordinarily the objectivity of a test is lower than its reliability due to the addition of a new cause of variation, namely, the difference in the personal equation of the different examiners. Some tests are more objective than others. A test of an individual's temperature, pulse, blood-pressure, finger-length, head-circumference, and the like, is usually much more objective than a test of his handsomeness or charm. Estimation of a man's height is rather subjective. The use of measuring instruments here as well as in education tends to increase objectivity. Tests are not totally subjective or totally objective. Objectivity, like reliability, is a matter of degree. Tests occupy points on a subjective-objective continuum with perhaps none located at either extreme. The degree of agreement in results secured by different examiners is the measure of a test's location on this subjective-objective scale.

Objectivity is an extremely important consideration in the selection of tests. So important is it that there is little exaggeration in stating that this criterion of objectivity is the mother of scientific educational measurement. For educational tests are an outgrowth of the extreme dissatisfaction with the subjectivity of previous methods of measuring the educational output. Progress in all sciences has been attended by a decrease in the personal equation through improvements in measuring instruments. *Verification* is the greatest word in the language of science. Education has been and still is to a large extent saturated with the personal equation. All progress in the development of education as a science is closely linked up with the creation of measuring instruments or measuring methods whose application yields verifiable results.

How may a test's objectivity be increased? The problem in education is no whit different from the problem in other sciences. The first step in its solution is to do everything possible toward

increasing the reliability of the test according to the methods sketched in the previous section. The second step is to determine, wherever possible, the amount and direction of the personal equation of the different examiners, and to allow for them. For some time to come improvement in reliability will be the most convenient and promising method of improving objectivity.

As with reliability, objectivity can be increased by a careful standardization of the entire testing process. If two examiners apply the test in different ways, disagreement is assured. If the method of scoring leaves room for the exercise of much judgment, disagreement is almost certain to arise. If there is a variation in the statistical method of computing pupil or class scores, it is hardly reasonable to expect results to agree. Adequate description can avoid most of the variation due to test application and statistical treatment. Much ingenuity is now being applied to developing completely objective means of measuring pupils.

CHAPTER IV

HOW TO SELECT AND CONSTRUCT TESTS— NORMS AND SCALES

Tests Should Have Norms Which Are Representative of the Group with Which Comparisons Are Desired.—There are in use two kinds of norms or standards which need to be distinguished, namely, standards *of* achievement and standards *for* achievement. The former means actual average achievements of age, grade, or other specified groups, whereas the latter refers to goals or objectives for these groups. The former is called *norms* and the latter called *standards*.

Norms are more valuable when they are representative of the group with whom it is most desirable to make comparisons. If but one norm could be had, all would agree that this should be the norm for all pupils in the country.

Tests Should Have Norms Which Are Stable.—Norms are more valuable when they are stable. The stability of a norm is a function of the satisfactoriness of the sampling and the number of cases. A hundred cases chosen carefully so they will be truly representative of the group for which a norm is being established are better than a thousand chosen with a bias, but when the sampling is equally well made the more cases there are the more stable the norm is, i.e., the less it will change with the addition of more and more cases.

Tests Should Have Norms Which Are Described in Detail and Reported in Full.—Test norms should have the method of their derivation clearly described. This appears self-evident, yet it is not at all uncommon to find a statement of norms without any explanation as to the method of their derivation.

Test norms should be reported in full. One author reports as norms for his test the highest score ever made in the test by any one class in the grade. This may have been done to stimulate teachers to special effort to bring their class up to this high standard. But such a stimulation is as liable to be unwholesome as beneficial since it may lead to overemphasis upon one sub-

ject. It is well to give the highest score and lowest score, or better the upper quartile and lower quartile scores, or, better still, all the percentile scores, for the fuller the norms are stated, the better. But whether several norms are given or only one norm, the best single score to report is some average measure.

Test norms should be both universal and local. An average norm for wide areas is useful but so are separate norms for a great many typical locations.

Test norms should be reported for both age and grade. Ballard complains that most of the norms developed in America are useless in England because our norms are grade norms. Age norms would be almost as valuable in England as in America. Usually nowadays age and grade norms are not given separately but are used to make age scales and grade scales and are imbedded in a table of age scores and grade scores respectively. Again, age norms permit the computation of reading age and reading quotient, spelling age and spelling quotient, and mental age and intelligence quotient. Numerous instances of the important functions which such measures serve have been illustrated many times throughout this book.

Tests Should Avoid Undistributed Scores Caused by the Tests Being Too Easy, Too Difficult, or Having Too Coarse Scaling or Truncated Scale Scores.—The fundamental aim of all testing is to reveal correct differences between pupils or groups of pupils. To reveal correct differences, a test must not only be valid but must possess, among others, the following characteristics:

1. Every pupil should make some score larger than zero. If every pupil makes a zero score it is utterly impossible to tell which is the best, average, and stupidest pupil. If only one pupil out of a class makes zero, there is no way to determine just how much more stupid he is than the rest of the pupils. Zero-score pupils are unmeasured. The range of ability in a class is usually very great, so, if the least able pupils are to make a score, the first elements of all difficulty tests must be within the ability of the least able and hence far easier than would be required for the abler pupils. The criterion requires that rate tests also must be composed of test elements whose difficulty is within the ability of all the pupils and which give a sufficiently long time limit. In an initial test in a Ph.D. research several

pupils in one of the experimental groups made zero scores. In the final test, some time after, they made scores above zero. The conditions of the research required that the amount of their improvement be known. How much did the pupils improve? Nobody knows. It may have been and probably was a small amount, but it may have been enormous.

2. No pupil should make a perfect score. Perfect-score pupils are unmeasured just as zero-pupils are unmeasured. In the case of perfect scores it is not known how much better the pupils are and in the case of zero scores it is not known how much worse they are.

3. There should be no undistributed scores, whatever. A test often yields undistributed scores when there is not a single zero or perfect score, and these may occur anywhere between the lowest and highest scores inclusive. These undistributed scores are produced by coarse scoring. The coarsest possible method of scoring is the "all or none" method. To score pupils on a test as either "passed" or "failed" is an example of the "all or none" method, and gives very undistributed scores, for so far as the scores indicate all who receive a pass are exactly alike.

How fine should the scoring for a test be? The fineness of the scoring depends upon the uses to be made of the results. The following, however, will serve as a rough general rule: *Select tests which will separate the pupils into at least seven groups of ability, and not less than thirteen if the data are to be used for correlation.* The above numbers, seven and thirteen, are minimum numbers. The finer the grouping the better. If the pupils are separated into less than seven groups of ability the results will have very limited uses, and if less than thirteen the influence of coarse scoring upon the coefficient of correlation will not be negligible. Among difficulty tests that one provides best against any sort of undistributed scores where the easiest test element is easy enough for all the pupils and where succeeding elements progressively increase in difficulty by small steps to a point beyond the ability of the ablest pupil. It is not necessary for the items to be arranged in order of difficulty if the time allowance is generous. A very fine scoring of a few test elements will, however, produce the same effect as increasing the number of test elements.

It was once deemed important that items should be arranged

in tests in the exact order of difficulty. Today this is not considered to be so essential except in tests where few pupils have time to finish. In this case it is better to provide particularly difficult items at the end of the test. These items would be a waste of time for slow pupils who generally are the less able ones, whereas they may be just the items needed to really differentiate among the very able pupils who get that far. Also they provide something to occupy the superior pupils until time is called, thus preventing them from finishing to the embarrassment and disturbance of slower pupils.

In the foregoing discussion of undistributed scores it has been assumed that each examiner will desire a score for each pupil, for unless such scores are secured a test cannot serve its most vital functions. In case only a class score is desired a few undistributed zero and perfect scores would do little or no harm if the median of pupil scores be the method of computing class scores. If, on the other hand, the mean of pupil scores is taken as the class score, undistributed extremes may seriously affect the size of the class score. Thus several recent tests make the computation of class means difficult or impossible by stopping their grade scores at 9.0. The simplest solution is to avoid such tests.

4. The test should be scaled and the standardized method of scoring should utilize these refinements of the scaling. The exactness of the scaling conditions the exactness with which differences between pupils can be measured.

5. A corollary of the preceding paragraph is that a test should yield a statistical result. All measurement in descriptive words should give place to mathematical statement. Supervisors, for example, frequently rate teachers without developing any statistical system of recording and combining their ratings. It is mainly in the realm of subjective estimates that non-statistical measuring occurs. Recently an experiment was undertaken to determine by means of standard tests just how accurately supervisors could estimate the efficiency of certain teaching methods. When the time came to compare test results with the judgment of the supervisors, no worthwhile computations could be made, for the supervisors had not kept any statistical records.

6. Finally, correct differences cannot be revealed unless the two scores yielded by each rate test be reducible to a common

denominator. Consider this situation from the *Courtis Addition Test*. Pupil A makes a speed score of 10 and an accuracy score of 90% while Pupil B makes a speed score of 12 and an accuracy score of 75%. Which pupil has made the better showing? As long as speed or accuracy is left free to fluctuate up and down in a sort of see-saw manner, no satisfactory comparisons between scores can be made unless a table is available for transmuting all scores to a constant speed or constant accuracy.

Perhaps the quickest method of determining the accuracy equivalence of a given amount of speed would be to adjust the weighting assigned until there has been secured the highest obtainable self-correlation between scores from two applications of the same rate test to the same pupils, when the scores correlated represent a combination score for both speed and accuracy.

Tests Should Yield Both Grade and Age Scores, or Other Appropriate Scale Scores.—There are numerous methods of scaling tests. First, there is the *goal scale* used by Courtis in connection with the *Courtis Supervisory Tests*. Any pupil whose score on the test falls between, say, 20 and 25 words spelled correctly on a particular spelling test is considered to have attained an appropriate spelling goal and is scored 1000. Any pupil who falls between, say, 17 and 20 is scored 500 and so on down to zero. Second, there is the *frequency-of-occurrence scale*. In the *Jones' Vocabulary Test* the score a pupil receives for knowing a certain word depends upon the frequency of that word's appearance in ten primers. In similar manner the degree of an individual's emotional aberration, as determined by the *Kent-Rosanoff Free Association Test*, is measured by the rarity of the individual's responses to the test.

Then there is the T scale used mostly in research, the percentile scale, popular in connection with tests for adults, and the age scale and grade scale, most useful in connection with tests for primary, elementary, and secondary schools. These scales will be described and their uses will be treated more fully later. Suffice it to say here that tests for levels below that of college should usually permit the determination of both age scores and grade scores. This proposition does not, of course, apply to teacher's examinations, although a plan is provided later in this book whereby such tests may yield grade scores or age scores.

CHAPTER V

HOW TO SELECT AND CONSTRUCT TESTS—SCORING

In test construction the prime requisite from the point of view of scoring is that those pupil reactions to the test which are to be scored be as simple, abbreviated, and controlled as possible, and that the reactions have a definite spatial location. With the exercise of some ingenuity, and when it is not important to diagnose the method of solution, a pupil's most complicated mental processes can be measured even when he reacts to each test element with no more than a word, a letter, a check, a number, or the like. The excellence of the pupil's solution of a long reasoning problem in arithmetic can be condensed into a few figures—the answer. If the pupil's reactions are simple and abbreviated they can be scored very rapidly and accurately, and with very little disagreement among the scorers.

Again, a test must also so control these reactions that only one kind of simple reaction will be correct. If any one of ten different words, or letters, or numbers is correct, scoring will be greatly slowed up and judgment must be more and more exercised and the net result is uneconomical, inaccurate, and subjective scoring. If only one reaction is correct for a given test element it is possible to make out a set of correct answers. These correct answers may be placed beside a pupil's answers, and then scoring becomes merely a matter of making simple, unthinking, visual comparisons.

Finally, the test must be so constructed as to give a definite spatial location to a pupil's answers. In any case this is a decided convenience; it is particularly so when a pupil's reactions all consist of a check mark or an underlining, where correctness depends not so much upon what is done as where it is done. Spatial location is secured by the provision of a square, circle, or other special place where the pupil is to make his mark. Consider, for example, how long it would require to announce the results of a presidential election if ballots did not spatially locate the voter's vote.

The problem of constructing a test so that scoring will be efficient is shown by the following evolution of an extract from a test for military aviators which the writer aided Thorndike in constructing. (Instructions are omitted.) Note first that the nature of the test question permits a long, qualified, unscorable answer. Note second that there is no prescribed place where the answer must be written. The following test element is a perfect illustration of what not to do.

Compare the lines as they were before with what they are now.

The test element is restated in better form below, though it is still inexcusable. Note that the nature of the test element encourages a briefer answer, and tends to control the type of answer.

Are the lines shorter than they were before, longer than they were before, or the same as they were before?

The test element is restated again in a still better form. The aviators were instructed to write the appropriate number in the parenthesis as I have done in the illustration. Note that the answer is simple, abbreviated, controlled, and located somewhat apart from the statement of the question.

Are the lines (1) shorter than they were before, (2) longer than they were before, or (3) the same as they were before? (2)

The above is the first form in which the question was actually stated. Note that a column of correct answers, properly spaced, could be placed beside a column of an aviator's answers in such a way that all errors could be detected with great accuracy and rapidity, and even more so if the answer were placed beyond the right edge of type.

But suppose the lines or trenches really are (2) i.e., longer than they were before. For the aviator to report to the Intelligence Officer that the lines are shorter than they were before is to make a more serious mistake than if he were to report that they are the same as they were before. The former should be penalized, say, two points and the latter only one point. Consider how the following re-arrangement facilitates the assign-

ment of the proper amount of penalty, if such partial scores are justified by the additional accuracy of measurement thereby secured.

Are the lines (1) shorter than they were before, (2) the same as they were before, or (3) longer than they were before? (3)

Since in this case the answer should be the lines really are longer than they were before, the correct answer is 3. If 2 is found in the parenthesis, it should be penalized 1 point. The difference between 3 and 2 is 1 point. If 1 is found in the parenthesis, it should be penalized 2 points. The difference between 3 and 1 is 2 points. Thus the test element has been so constructed that the difference between the correct number and the number appearing in the parenthesis gives instantly the proper amount of penalty. Without such simplification of scoring the extensive use of mental tests in the military service during the war would not have been possible, nor would there be great promise for their future use in education.

Below are extracts from a variety of tests, which illustrate how not only tests but ordinary examinations can be so constructed as enormously to reduce the inaccuracy, subjectivity, and time of scoring.

EXTRACT FROM RUGER'S PROVERBS TEST

DIRECTIONS: *In column No. 1 write opposite each English proverb the number of the African proverb which most nearly means the same thing as the English proverb (see below for African proverbs). (Do not write any number twice—omit no number—write only one number opposite each letter.)*

Column		ENGLISH PROVERBS
1	2	
<div style="border: 1px solid black; width: 80px; height: 80px; margin: 2px;"></div>		a. First catch your hare.
		b. Curses come home to roost.
		c. Milk for babes.

AFRICAN PROVERBS

1. Ashes fly in the face of him who throws them.
2. I nearly killed the bird. No one can eat nearly in a stew.
3. If the stomach is not strong, do not eat cockroaches.

EXTRACT FROM THORNDIKE'S MENTAL ALERTNESS TEST

Make a cross in the square before the best answer to each question.

- | | | | | | | | | | |
|--|---|--|--|--|--|--|--|--|--|
| <p>1. Why are prunes a good food?
Because they</p> <table border="1" style="display: inline-table; vertical-align: top;"> <tr><td style="height: 20px; width: 20px;"></td></tr> <tr><td style="height: 20px; width: 20px;"></td></tr> <tr><td style="height: 20px; width: 20px;"></td></tr> <tr><td style="height: 20px; width: 20px;"></td></tr> </table> | | | | | <p>4. When you feel that affairs in
your town are badly man-
aged, should you?</p> <table border="1" style="display: inline-table; vertical-align: top;"> <tr><td style="height: 20px; width: 20px;"></td></tr> <tr><td style="height: 20px; width: 20px;"></td></tr> <tr><td style="height: 20px; width: 20px;"></td></tr> <tr><td style="height: 20px; width: 20px;"></td></tr> </table> | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| <p>grow in California
are wholesome and eco-
nomical
are served in boarding
houses
make an attractive dish</p> | <p>do nothing at all
growl to your friends
get out and work to have
things changed
go to church</p> | | | | | | | | |

EXTRACT FROM PRESSEY'S MENTAL SURVEY TEST ¹

X. Analogies.

girl—woman: boy—man
Examples: sun—day: moon
good—bad: big

- | | |
|--------------------------------|---|
| 1. woman—girl: man | 11. hill—valley: high |
| 2. kitten—cat: puppy | 12. arm—elbow: leg. |
| 3. sky—blue: grass | 13. truth—falsehood:
straight line |

EXTRACT FROM GREENE'S ORGANIZATION TEST ²

Write numbers in these spaces

- | | |
|-------------------------------------|--|
| (1) (2) (3) | |
| 1. a dog, a boy, had . | <table border="1" style="width: 150px; height: 20px;"></table> |
| (1) (2) (3) | |
| 2. of the cold, afraid, they were . | <table border="1" style="width: 150px; height: 20px;"></table> |
| (1) (2) (3) | |
| 3. I am, see, how tall . | <table border="1" style="width: 150px; height: 20px;"></table> |

EXTRACT ³ FROM OTIS' GROUP INTELLIGENCE SCALE

MEMORY

DIRECTIONS: *Read each question and if the right answer, according to the story, is YES draw a line under the word YES. If the right answer is NO, draw a line under the word NO. But if you do not know the right answer, because the story didn't say, draw a line under the words DIDN'T SAY.*

¹ Issued by Pressey, S. L. and L. W., University of Indiana, Bloomington, Ind.

² Issued by Curtis, S. A., University of Michigan, Ann Arbor, Mich.

³ Copyrighted 1919 by World Book Company, Yonkers-on-Hudson, New York.
Used by permission of publishers.

Sample:	Was the story about a king?	yes no didn't say
	Was the king's daughter sixteen years old?	yes no <u>didn't say</u>
	Was she ugly?	yes no <u>didn't say</u>

Begin here:

1. Was the king fond of hearing stories? (yes no didn't say) 1.
2. Did the king offer his daughter to any one who could tell him a story that would last forever? (yes no didn't say) 2.
3. Did he offer all his kingdom also? (yes no didn't say) 3.
4. Did he say, "but if he fails he shall be cast into prison"? (yes no didn't say) 4.

Test Should Permit the Use of Scoring Devices.—Since scoring is greatly facilitated by mechanical scoring devices and since the possibility of employing such devices is dependent upon the form of arrangement of the test material, a brief discussion of these devices is pertinent at this point.

There are many forms of these mechanical devices depending upon the form of the test which they are designed to score. When all the pupils' answers are written at a definite place on the right or left edge of the test sheet a convenient device is a printed scoring stencil or a test sheet which has been correctly filled out by the scorer. The key sheet can be so superimposed on the pupil's sheet that nothing but the pupil's column of answers shows immediately beside the correct answers.

Again, there are tests of such a nature that what the pupil does is relatively insignificant but where he does it is all-important. Such are tests where the pupil is instructed to underscore the appropriate word, or check the appropriate reason, or cancel the appropriate letter, etc. The scoring device already described may be used to advantage in this situation, but some form of transparent sheet frequently works better. Celluloid or any kind of transparent material may be placed over a correctly filled test, and a dot can be made on the celluloid sheet just over the place which is correct. The transparent sheet may then be used for scoring the pupils' answers. Otis makes an extensive use of just such a device for scoring his group intelligence test.

Finally, if the test is so constructed that scoring will be facilitated by making all of the pupil's test sheet invisible ex-

cept the spot where the correct answers should be, small apertures may be cut through a blank sheet at such places that only the correct-answer spots will be visible. The same result may be secured by placing a sheet of celluloid over a test sheet and by so painting the celluloid with black paint that nothing but the desired spots will be visible. These perforated scoring devices may also be used to facilitate the counting separately of items mixed in one text. The *Mixed Fundamentals of Arithmetic Test*,¹ has addition, subtraction, multiplication, and division examples so mingled on one test sheet that the pupil is frequently forced to shift his processes, and often to decide by the nature of the sign just what sort of an example it is. In the instructions which accompany this test, it is suggested that the computation of a separate score for each fundamental, if desired, may be facilitated by perforating four fresh test sheets. The first sheet should be so perforated that when placed over the pupil's test only addition examples are visible. The second sheet should make visible only subtraction examples; and multiplication and division should be treated similarly. It is rather unsafe to use this perforated scoring device for determining whether a pupil has made a mark at the right spot, for he may have made two marks—at the right spot visible through the aperture and the wrong spot hidden by the scoring stencil.

Clapp and Young (*Self-Marking Tests*, Ginn and Co.) use carbon on the back of a test sheet and under the correct answer only. When a pupil marks the correct answer his mark is transmitted by the carbon to a blank sheet under the test. When an incorrect answer is marked, no mark is transmitted. The pupil's score is then determined by counting the number of marks on the blank sheet.

Peterson and Peterson (Lincoln School Supply Company, Lincoln, Nebraska) have each pupil record his answer on a single separate answer sheet. The first row on the answer sheet looks like this.

1. *a* *b* *c* *d*

The pupil looks at the true-false or multiple-choice test item number 1. If he thinks the first choice is the right answer, he circles *a*, if the second, *b*, if the third, *c*, and if the fourth, *d*, and

¹ Bureau of Publications, Teachers College, Columbia University, New York.

similarly for other test items. The answer sheets are then stacked and firmly anchored in a simple machine by the scorer. An awl is then used to make a hole through answer sheets at *b*, if *b* is the correct answer to test item 1, and so on for the other items. The pupil gets one point of score wherever his circle is found to be around the awl hole. They call their device a *Perfo-Score*.

Roberts (Educational Machine Company, Point Marion, Pa.) has designed a more elaborate machine called *Krexite* which, after setting, punches all holes with a turn of the crank.

Stenquist (Bureau of Educational Research, Baltimore) places a colored spot on the correct answers by means of a mimeograph machine, using hand feed. One paper is marked at a time.

Peterson and Peterson also have a *Thermo-Score*. A batch of answer sheets, after being marked by the pupils, are gently heated and the invisible "magic ink" on the correct answers shows spots of color. They also have a *Chemo-Score*. The pupil uses a moistened brush to mark the answers. When he is right, green instantly shows. When he is wrong, red appears.

Other workers whose names have become prominent in this effort to take the labor out of scoring and to enable "morons to measure geniuses" are Otis, Toops, Pressey, Johnson, Wood, and Kryzanowsky.

The International Business Machine Corporation, New York, has gone a step further. Its machine (for rent only) both scores and counts the score. The pupil marks the answer sheet. Answer sheets are fed into an electrical machine. The carbon in the pupil's pencil mark, when made on the correct answer, interrupts a faint electric current. The machine adds one point for each interruption.

Group tests in particular make objective scoring imperative. The nature of certain tests and the illiteracy of young pupils has required individual testing, i.e., the testing of one pupil at a time. The nature of other tests and the literacy of older pupils permits group testing, i.e., the testing of many pupils.

A heated controversy has been going on concerning the advantages and disadvantages of each method of testing, and this controversy continues in spite of the fact that skillful test constructors have now adapted almost all varieties of tests to permit group testing of illiterates.

Even when group testing is feasible, it is claimed that a more accurate diagnosis can be made when each pupil is tested individually. This claim is based upon the assumption, first, that the appearances and incidental reactions of a pupil are valuable indices of his special defects or special strengths and that these indices are observed better during an individual examination. The second assumption is that the examiner can better select for each pupil those tests which will reveal significant symptoms, for it often happens that some reaction on the part of the pupil will give the examiner a "lead" which it is highly desirable to follow up. Such rapid adaptations are manifestly impossible in group testing. Finally, some examiners hold that testing conditions can be more carefully standardized by individual testing. Early psychological investigators considered themselves unusually virtuous when they took time to administer all tests individually "with special care," as they said.

Group measurement is enormously economical in time. To administer a thirty-minute individual test to a group of 500 pupils would, when all wastage is counted, take about 300 hours of the examiner's time, whereas, under certain circumstances, a thirty-minute group test could be administered to all pupils in about forty minutes. Even though the 500 pupils were tested in groups of only fifty, a great saving of time would be effected. It is this great expense in time that has delayed educational measurement in the kindergarten and primary grades. The economy of group testing is further illustrated by the psychological examination of soldiers during the war. Several tests were given to many hundreds of thousands of soldiers. Each test could have been administered individually to each recruit. To have done so with the staff available would have required all the years of the war, when speed was imperative. Substantially the same situation confronts those who are introducing measurement into education. It is useless to attempt the measurement of millions of pupils with individual tests. To a very large extent educational measurement must be group measurement.

Group testing may, under certain conditions, be fairer to the pupils tested. In experimentation it is often important to know the amount of change made by each pupil in a class during two weeks. It might take a single examiner a week to test every

child by the individual method. The last pupils tested would thus have an extra week's advantage if learning were being measured, or a week's disadvantage if forgetting were being measured. Again, a test is often of such a nature that one pupil can partially prepare another. The first pupils tested can then spread information through the entire class or school. Finally, for some tests, it is especially difficult to standardize the personal equation of the examiner. Such a variable operates to the advantage of some pupils and to the disadvantage of others. Group testing makes the personal equation more nearly constant for all pupils within the group.

What then is the conclusion of the whole matter? Individual testing and group testing each secures special values. The method adopted in the psychological examination of soldiers will probably come into common use in all educational measurement whether done for purely pedagogical or clinical purposes. The initial tests given the soldiers were group tests. These revealed the illiterates and those who were in some way abnormal. The illiterate and abnormal groups were then intensively measured with individual tests. The diagnoses afforded by the group tests were accepted for the vast majority of the recruits. In time, school psychologists will not wait until abnormal cases are sent to them for diagnosis. They will sweep through the schools with a net of group tests and catch their own cases for intensive study. Even for the special cases, what with the development of group tests for illiterates, it is worth considering whether the greater number of group tests which may be given within an equal time-interval may not give a better diagnosis than the fewer individual tests. A good practical rule is to *first give group tests, accept their diagnosis for most of the pupils and give further group or individual tests to the few pupils, who, according to the group tests, need special study.*

Tests Should Eliminate Additions to the Score Due to Chance When Such Is Present and Elimination Is Desired.—Sometimes, but very rarely, it is important to know approximately how many test items were actually known by a pupil. This requires that we strip from the total number right those items answered correctly by sheer guess. It is also important to eliminate guessing in tests which have time limits so short that few pupils are able to finish the test, and in which rapid guessing

can make large additions to the score. To illustrate with the sample test presented in Chapter II:

Number of correct underlinings = 75

Number of incorrect underlinings = 15

Number of omissions = 10

(A) Pupil's score = number correct - number wrong.

Pupil's score = $75 - 15 = 60$

Let us consider first the reason for expressing a pupil's score as the number correct minus the number wrong. Imagine a pupil who is absolutely innocent of any knowledge of the physical features of the United States. Were such a pupil to take the above test and were he to mark every statement he would according to the theory of chance mark 50 statements correctly and 50 incorrectly. The chances of his guessing right or wrong are fifty-fifty or one to one. His score on the above test would be:

$$\text{Score} = 50 - 50 = 0$$

In short, the pupil's knowledge is zero and the method of computing his score gives him zero. Suppose instead that he knows 60 statements and guesses at the other 40. Of the 40 guessed at he would, according to chance, get 20 correct and 20 wrong. That is, even though his real knowledge is 60 he will show 80 correct ($60 + 20$) and 20 incorrect. The method of computing his score brings out his real knowledge.

$$\text{Score} = 80 - 20 = 60$$

A pupil who marks every statement correctly makes a perfect score, as follows:

$$\text{Score} = 100 - 0 = 100$$

Observe that no account is taken of omissions. Only the corrects and incorrects figure in the pupil's score. When the time allowed the pupils to take the test is made short in order to test each pupil's speed of work there will, of course, be many papers showing several omissions each. In all such cases omissions should be ignored, just as we have done above, in computing scores. Even when the time allowed for the test is ample for each pupil to mark every statement, there will still be an occasional instance of omission due to carelessness or misunderstanding of instructions or a puritanic conscience against

increasing the score by gamble guess-work even when the instructions urge guessing.

When the time is ample for even the slowest pupils and when all are instructed to mark every statement, it is much more convenient to compute a pupil's score according to the formula which follows:

Score = (number of statements) minus 2 (number marked incorrectly)

The formula for eliminating guessing from the score when there are three choices, i.e., one right answer and two wrong answers is

$$\text{Score} = R \text{ minus } \frac{1}{2}W$$

When there are four choices the formula becomes

$$\text{Score} = R \text{ minus } \frac{1}{3}W$$

Thus the generalized formula is

$$\text{Score} = R \text{ minus } \frac{1}{N-1} W$$

when N is the number of choices.

If a pupil is asked to name the opposite of *hot*, he may say *cold* or any other word in his vocabulary. Here we have a test item with one right answer and, say, 10,000 wrong answers. The formula becomes

$$\text{Score} = R \text{ minus } \frac{1}{10,000-1} W = R \text{ minus } \frac{1}{9999} W$$

Thus, for all practical purposes the formula becomes

$$\text{Score} = R$$

when the choices are numerous, as they are in all *recall* examinations.

Guessing may likewise be eliminated, though less easily, from other forms of objective test, as for example, from matching tests such as one which appears in the *Health Awareness Test*, Bureau of Publications, Teachers College, where items *a, b, c, d*, etc., must be matched correctly with items *1, 2, 3, 4*, etc. Zubin¹ has developed the proper formula.

¹ Zubin, Joseph, "The Chance Element in Matching Tests," *The Journal of Educational Psychology*, December, 1933.

Tests Should Achieve the Desired Weightings for Different Tests and the Various Portions of the Same Test.—If an examination is divided into, say, two portions, and it is desired that the two portions be weighted equally in the total score, and it is desired to approximate this without recourse to complicated statistical operations afterward, the best plan is to have the same number of items in each part and have them represent about the same spread of difficulty.

If it is desired that the second part have, say, only one-third the weight of the first part, the second part should contain one-third as many items as the first part and should represent about the same spread of difficulty.

The same principle holds for different tests as for parts of one test.

Two points are worth noting in scoring essay tests. The test should be scored in terms of points, either one point per item or a varying number of points depending on the weighting it is desired to give the various items. Second, it is better to score the first item for all the pupils, and then the second item for all the pupils, and so on.

Tests Should Permit of Accurate and Economical Scoring and of Pupil Scoring.—Multiple-choice examinations permit of much more economy in scoring. If a copy of the test has been marked by each pupil, the teacher can take an unused test sheet, fill it out correctly, lay the correct column of answers beside each pupil's column of answers, and quickly mark whether the pupil's answers are correct or incorrect. If a copy of the test has not been placed in the hands of each pupil, but each has instead written *True* or *False*, or made a check or cross after the number of each statement, the teacher can take a page of paper similar to that on which each pupil has indicated his answers, copy the numbers just as they are and just as they are spaced on the pupils' papers, write after each number the correct answer to the statement of that number, place this column of correct answers beside the column of pupil answers and mark those which are correct and incorrect. This last scoring method presupposes that pupils have used ruled paper, and that each has written his numbers in a vertical column according to a particular spacing recommended by the teacher. Last and best, each pupil can score his own or his neighbor's paper.

If the method of pupil scoring is adopted, the teacher should read the correct answers while the pupil checks his own. If the pupil does not have a copy of the statements before him, the teacher should read each statement before giving the correct answers, in order that the pupil may know what statements he got correct or incorrect. When every pupil's answers have been marked and when his score has been computed and recorded on his examination paper, the teacher should ask all the pupils who missed statement number 1 to hold up their hands, and then all pupils who missed number 2 to hold up their hands, and so on. The teacher should make a record of the number of pupils missing each statement, and then collect all papers.

The fact that pupil scoring will relieve the teacher of much obnoxious drudgery, does not justify the inference frequently made that what is non-educative drudgery for the teacher will also be non-educative drudgery for the pupils. On the contrary the most favorable teaching opportunity that ever comes to a teacher is the period immediately following an examination. The pupil's interest to know what parts of the examination he missed and what he got correct is then at white heat. Witness the interested discussion among pupils immediately following an examination. It is inexcusable neglect of an educational opportunity not to capitalize these precious moments for correcting erroneous ideas, clinching right ideas, and filling up mental spaces where ideas are not. These values can best be realized by having pupil scoring and by stopping to discuss points where pupils have trouble. Of course not every correct answer indicates knowledge, but the pupil himself usually knows when he knows.

The multiple-choice examination is also more educative, because it is likely to be given more frequently. The experience of Kirby, Courtis, and others with practice tests shows that a pupil learns more during testing periods than during teaching periods. We really teach when we test. This examination covering as it can a wide range is an ideal method of review. It reveals to the pupils just where their difficulties lie. Testing is one of the best ways of teaching.

With a method of testing available which involves no drudgery to anyone, testing is likely to be more frequent, and this means more complete and timely information about the abilities

and difficulties of the various pupils, and about the successes and failures of teaching efforts. It has already been suggested that the teacher keep a record of the number or per cent of pupils missing each statement in the examination. This record will show what things have been well learned or poorly learned and well taught or poorly taught. Also it is a good thing for a teacher to check her own efficiency in general. This can be done by finding the average of the scores of all the pupils and by comparing this average with the total number of statements in the examination or at least the total number of facts the teacher has really attempted to teach the pupils. If the average score, corrected for guessing, is 20 out of a possible 40, the teacher's efficiency is 50%. Most teachers will be chagrined to find, if they use truly representative items in their examination, that their efficiency is below 50%. Similarly, a pupil's efficiency may be determined by the per cent of statements he got correct out of the total number of statements the teacher has a right to expect him to get. Before the examination is given the teacher should decide what items she has a right to expect the pupils to get correct. This same number should then be used for computing both pupil and teacher efficiency.

But pupils will cheat. To be sure some will cheat. It will advantage us nothing to delude ourselves into the belief that cheating will not occur. To do so would be to join the peerage of the ostrich that is fallaciously reported to stick its head into the sand and think itself safe, or of the partridge which dives into a snow bank and feels as secure of its safety as the hunter feels of his game. It would advantage us still less to compel honesty by so arranging all educational situations that there is no opportunity to be dishonest. The chances that the world will be so tender of a pupil's weakness are very few indeed. If a pupil has it in him to be dishonest, it is a genuine kindness for the teacher to find it out. The benevolent birch removes less epidermis than the rod of the law.

CHAPTER VI

HOW TO SELECT AND CONSTRUCT TESTS— INSTRUCTIONS

1. **Test Instructions Should Be as Brief as Is Consistent with an Adequate Understanding of What Is to Be Done.**—Besides consuming time, inordinately long instructions tend to produce confusion in the minds of the pupils. Even adults find difficulty in following through complicated instructions. It has been demonstrated frequently that even among so intelligent a group as school teachers there are always a few who cannot follow very simple directions. Long instructions so tax the memories of pupils that absolute essentials are frequently forgotten. To forget a single one of these essentials may markedly alter the child's score. Brevity is frequently sacrificed to pure irrelevancies. It is well to remember that the primary function of instructions is to give a pupil adequate, but not necessarily complete, information about the test. Their primary function is not to give the pupil a general education. To quote a remark by Patterson, "Test! Don't teach!"

Again, the longer we make the instructions, the more we add to the confusion of inexperienced examiners. The novice is never quite sure of himself unless the instructions are sufficiently brief that his memory span can embrace not only every step of the process, but also the proper sequence of the steps. The untrained examiner cannot give his sole attention to instructions. He must maintain order among a roomful of naturally disorderly creatures, keep track of his watch, handle the test sheets, see that preceding instructions are being followed, and the like. It is a real kindness to both examiner and pupils to make instructions no longer than is necessary.

But inadequate instructions are as bad as or worse than instructions which are too long. Inadequate instructions may wholly defeat the purpose of the test, or precipitate an avalanche of questions from the pupils. Instructions cannot be cut out of whole cloth. It requires both forethought and experimen-

tation to produce instructions which will cause the pupils to do just what is wanted of them, and which will anticipate questions by the pupils.

The omission of some points would be more disastrous than others. What the essential key points are depends, of course, upon the test. In the Thorndike *Vocabulary Scale*, for example, it is especially important that pupils be warned not to skip any words by accident. This is because the statistical method of computing scores for this test treats accidental omissions as though they were errors, and weights them very heavily. Below are a few quotations from existing test instructions which are key points.

As soon as you complete the first sheet, hold up your hand, and I'll give you a second one.

Read as rapidly as you can to still understand what it says.

Don't read anything over again.

You will have just one minute.

This is an addition test.

Check each sum before passing to the next example.

When I call 'stop,' draw a circle around the last word read.

You will be asked to reproduce from memory what you have read.

Your score will be the number of examples you get right.

You will be marked on both speed and quality.

Write your name and grade.

Some key points are so obvious that they will be recognized by anyone. Some are so subtle that only the intuitive or trained examiner can detect them. In sum, instructions should be as brief as possible, as adequate as is essential, and always consistent with the subsequent uses of results.

2. Test Instructions Should Employ a Demonstration and Preliminary Test.—An ounce of demonstration is worth a pound of words! It takes more words to describe effectively what is to be done than it takes moves to show what is to be done. Anyone can try for himself an experiment to discover whether it is easier to show than to tell. Probably due to primordial practice, children, not to mention adults, can imitate better than they can comprehend and follow linguistic directions. To accompany description with a demonstration not only caters to pupils who may get impressions easier through the eye or through the ear, but, what is more important, it gives to all an impression through both eye and ear. Demonstration has the still further

advantage of securing better attention, especially from the young children.

The demonstration may take any of several forms. In one test the examiner writes a sample test element on the black-board and works it out for the pupils just as they are to work out similar tasks contained in the test. But in most tests which employ the demonstration method, sample test elements correctly completed are printed on the test sheet. Here is an example of instructions for a test accompanied by such a completed sample:

"This is a test of common sense. Below are sixteen questions. Three answers are given to each question. You are to look at the answers carefully; then make a cross in the square before the best answer to each question, as in the sample:

SAMPLE	{	Why do we use stoves? Because
		<input type="checkbox"/> they look well
		<input checked="" type="checkbox"/> they keep us warm
		<input type="checkbox"/> they are black

"Here the second answer is the best one and is marked with a cross. Begin with No. 1 and keep on until time is called."

Thorndike has devised a novel test. This test attains the maximum of showing and the minimum of linguistic directions. So much is this the case that it may well be called a *pantomime* test. The whole test can be given without the reading or the speaking of a word by anyone. The test was devised, in fact, to measure the intelligence of army recruits who were illiterate Americans and immigrants who did not even understand spoken English. The recruits were given a test sheet containing diagrams, pictures, etc. The examiner placed before the recruits an enlarged form of the test which was similar to, but not identical with, the test in the hands of the recruits. The examiner did the enlarged test with a heavy crayon. The examiner's movements showed the recruits what they were to do with their own test sheet. This is a most ingenious test, but, when there is a common medium of communication, the best method of giving instructions is not by demonstration alone, nor by linguistic description alone, but by a happy combination of both.

When instructions are at all complex, they should, as a rule, be accompanied by a preliminary test. Even though every possible precaution be taken to make all pupils understand just what they are to do, one can never be quite sure that all do understand unless a preliminary test is given. A preliminary test has the additional advantage that pupils can make most of their test adjustments before beginning the test proper. Due to differences in nervousness, intelligence, etc., some pupils adjust quickly and some slowly. If there is no preliminary test, and if the time for the test is relatively brief, the rate of adjustment may materially influence the score, even when we are usually not primarily concerned with the measurement of this factor. The preliminary test should typify the nature of the test elements proper.

This preliminary test may be presented in various ways. Sometimes the examiner writes one or more typical test elements on the blackboard and the children do them more or less in concert. Obviously this method does not give the examiner a sure guarantee that each pupil understands what is expected of him.

A second method is to give each pupil an easy miniature test. The examiner can then go about the room and observe whether each pupil shows an understanding of instructions. The examiner can help any pupil do the first element or two if he does not understand. If this does not suffice, the pupil can be assumed to be incapable of doing the test at all.

A third method is to print the preliminary test on the back of the regular test sheet along with the instructions or to reserve the front page of a booklet for instructions and preliminary test. This method is most satisfactory of all. Its use is not universal because of the greater expense involved in printing on both sides of a test sheet or making a booklet.

A fourth method is a little less satisfactory and, as a compensation, less expensive. The instructions, demonstrations, and preliminary practice test can be printed on the same side of the sheet as the regular test, but clearly separated from the regular test. Pupils can be instructed to do the practice test, but not to begin the regular test until their work on the preliminary test has been inspected and they have received the signal to start the test proper. It is difficult to prevent a premature *mental*

start. If the test is a rate test such a premature start may be a serious factor.

A fifth method has been used. When the time element is not important, the elements of the preliminary test may be, so far as the pupil is informed, the first few elements of the regular test. After the test has been started the examiner can go about the room and give any needed help on the preliminary elements. In this case the preliminary elements will not be counted in determining the pupils' scores.

Sometimes practically all the advantages of all the methods can be secured by folding back the preliminary portion of the test in such a way as to conceal the regular test while the preliminary test is visible. This permits printing the test by a single impression, and thus reduces expense. If expense is not, however, a consideration, the folder or booklet test, with the entire front page exclusively reserved for name, grade, age, instructions, and preliminary test, is preferable.

3. Test Instructions Should Be Adapted to and Uniform for All Who Are to Be Tested.—How much adaptation is essential? In the testing of school abilities, the instructions for the test should be so simple that all may understand them. The instructions should be such that no child will fail to make a score just because he failed to comprehend the instructions.

How much uniformity is essential? Instructions contain mechanical and non-mechanical features. The mechanical phase has to do with getting the pupil's name, sex, age, grade, etc. Uniformity is not necessary because the important thing is to get this data of identification, even though it is necessary for the examiner to so vary the procedure as to write the pupil's name for him. The mechanical features do not assist the pupil with the test proper.

The non-mechanical features do determine to a certain extent, and frequently to a large extent, the score a pupil will make. It is far more convenient if these instructions are uniform from grade to grade. To cite one illustration, tests are frequently used in rural schools where several grades and many ages are grouped in one room. An examiner can test all these pupils at once if the instructions are uniform. Hence it is best for instructions to be both adapted to and uniform for all the pupils in all the grades.

The intelligence examiner will grumble because I have not been even more enthusiastic for absolute uniformity. The intelligence examiner frequently has only a minor interest in knowing whether failure on the part of the pupil is due to lack of comprehension of the instructions or due to the inability to do the test elements. His primary interest is to find out whether the child possesses sufficient intelligence to deal with the total situation. And therefore the measurer of general intelligence may be right in contending that instructions should be absolutely uniform for all ages. Otherwise the total situation would not remain constant.

But it is unwise to carry over to educational measurement a theory which is inapplicable. When an educator gives a vocabulary test, he is, as a rule, primarily interested to know what the pupil's vocabulary is, and only incidentally interested to determine whether the pupil possesses sufficient general intelligence to understand the instructions or overcome the mechanical difficulties of the form of the test. If a teacher measures her pupils' ability to add, she wants to know how well her children can *add*. She is not then interested in knowing how well they can understand her directions or read printed instructions. She wishes to reduce these irrelevancies to a minimum. Only in a test of reading ability is it perfectly legitimate to make the instructions an integral part of the test itself. Nor is this primary interest peculiar to education. Many psychological tests which are designed primarily to measure intelligence prefer to measure it by means of the test material rather than by the instructions.

If the above distinction is sound it is legitimate to construct different instructions according to the age and ability of the pupils, provided whatever instructions are used give in every grade an adequate understanding of what is to be done, which means that if sixth-grade instructions are more difficult than third-grade instructions, the former must still be easy enough for each sixth-grade pupil to understand what he is to do. In essence this means that in educational measurement adaptation has priority over uniformity. My thesis required both adaptation and uniformity because I think it is possible to secure both at once.

But it is frequently contended that there is no possibility of

securing adequate adaptation together with uniformity. It is claimed that the two characteristics are mutually antagonistic and that we cannot have our cake and eat it too.

It is held by some that words which are appropriate for third-grade pupils would insult eighth-grade pupils and words appropriate for eighth-grade pupils would be beyond the comprehension of younger pupils. It may easily be doubted that third-grade children appreciate "baby talk" as much as is claimed. Nor is it impossible to find words sufficiently simple that younger pupils will understand them and at the same time so dignified that older pupils will not resent them.

When a test is being selected for wide use throughout the country special care should be taken to see that instructions can really be kept uniform and yet be universally adequate and universally just. In the first place instructions should not require for their proper presentation material which some places may not have. Instructions should not require, for example, a blackboard unless a blackboard is likely to be available wherever the test is to be given. Again, the instructions should employ neither words nor illustrations which have local significance only. When Woody could not find a universal term in use which meant an *addition* example, he secured universality by giving other terms in common use and suggested that examiners use the terms current in the grade or locality where the testing is being done. Again, an examiner once discovered that the standard instructions lacked sufficient universality because they failed to take into consideration the fact that some pupils are left-handed. Illustration of elements conditioning universality could be multiplied.

4. The Order of Test Instructions Should Be the Order of Doing.—It is probable that pupils can carry out instructions with greater ease when the order of the instructions is the order of doing. Long instructions are far more tolerable when the steps in the direction come in the same order as the steps of the process the pupil must go through. The demonstration is easier to imitate when the pupil does not find it necessary to transpose, in the process of doing the test, the steps observed in the demonstration.

5. Test Instructions Should Be Broken into Action Units.—The strain upon the pupil's memory is not nearly so great when

the instructions are broken into action units. Wherever possible the pupil should carry out one direction before any other directions are given. The instructions which follow are not broken into action units.

The experimenter holds the sheet before the class and says: "This sheet contains some incomplete sentences, which form a scale. This scale is to measure how carefully and rapidly you can think and especially how good you are in your language work.

"You are to write one word on each blank, in each case selecting the word which makes the most sensible statement.

"You may have thirty minutes in which to sign your name at the top of the page and write the words that are missing. The papers will be passed to you face downward. Do not turn them over until we are all ready. After the signal is given to start, remember that you are to write just *one* word on each blank and that your score depends on the number of perfect sentences you have at the end of thirty minutes."

It is easy to imagine just how little a pupil would remember of the key points in the latter set of instructions after the excitement of passing papers, writing names, and the like. When the order of instructions is the order of doing, and when the instructions are properly segmented by action, the instructions intimately concerned with each step of what the pupil is to do immediately precede that step. The pupil can give his undivided attention to that particular bit of instruction. When this principle is not satisfied the pupil is trying to grasp what is coming next and at the same time trying frantically to hold on lest what he has already heard escapes.

6. Test Instructions Should Equalize Interest.—There are numerous factors besides interest which condition ability. Interest is dignified with special consideration because of its large effect upon the pupil's score. Interest determines effort. A pupil with high ability may show a range of interest from zero to high intensity, and hence a similar range of effort.

Shall standardization be upon a high plane of interest or upon a low plane? And how shall the desired stratification be secured? Experimental results have not yet shown whether it is easier to equate interest on a low plane, medium plane, or high plane. Hence general common-sense experience must decide. Practical considerations rule out the offering of rewards high enough to secure the intensest possible interest. Normal

life interests vary so greatly that they cannot be taken as a criterion. The fact that tests are not so educative when taken with low interest as when taken with high interest tends to rule out attempting an equalization of interest on a low plane. Furthermore, performance on one test does not seem to agree so well with performance upon a duplicate test when interest is on a low plane. In the absence of reliable evidence, the best guess is that performance is more constant and is a better index of the ability being measured when interest is at the maximum attainable by practicable methods.

What motivation can be legitimately employed? Unless such will defeat the object of the test, the pupil should be informed of the general purpose of the test and when it is not perfectly obvious, of the general method by which he is to be scored. A pupil will be more interested who is told that the purpose of the test is to discover how rapidly he can read and then how accurately he can answer from memory questions upon what he has read, and hence his score will depend upon the number of seconds required to read a passage and the number of questions he can answer correctly upon what he has read. A detailed discussion of the purposes and methods of the test should not be attempted because of the necessity for brevity, and sometimes because of a necessity for concealing from the child the exact method of scoring. Secrecy is occasionally necessary in experimental work and in cases where the score is at the mercy of the pupil's honesty or lack of honesty.

The behavior of the child and the testimony of adults bear eloquent witness to the potency of rivalry as a begetter of interest. Probably no stimulus at the disposal of the school is so powerful, natural, and generally healthful.

It is scarcely necessary to point out, however, that it will soon become impossible to secure interest through any method unless the pupils have an opportunity to learn how well they did on the test.

Some may think that the device of securing interest by means of some form of rivalry is artificial. We cannot be sure of this. Most of the games voluntarily selected by children and adults would never be selected for their own sake. With children as well as adults rivalry is itself intrinsically satisfying. Remove the contest feature and how long would men and women lay

card on card, or men punch ivory balls into holes with a long slender stick, or would war even remain the engaging pursuit that it is? Interest through projects is excellent, but interest through rivalry is not always artificial.

7. Test Instructions to Pupils Should Be Accompanied by Instructions to Examiners.—Instructions to pupils should be accompanied by instructions to examiners telling how the test is to be applied, because it is a question which needs instructions more, pupil or examiner. Instructions to the examiner should be in steps easy to comprehend and follow. This easy use can be facilitated in two ways. First, the author of the instructions should formulate for the examiner the exact words to say to the pupils and insert between various units of directions to pupils, the necessary directions to the examiner. And, secondly, when the instructions to the examiner are inserted among instructions to pupils, the latter should be set off from the former in some convenient fashion. This can be done by numbering, paragraphing, underscoring, or italicizing the words to be said to the pupils.

Better still, most of the instructions for the teacher should be incorporated into the instructions for the pupils so that everyone will help keep everyone else from forgetting them.

8. Tests Should Permit Administration without Undue Inconvenience.—The methods of applying essay examinations are too well known to require comment, so the discussion will be confined to multiple-choice tests. The best way is to print, mimeograph, or otherwise duplicate, the examination, and place a copy in the hands of each pupil. But there are numerous schools which lack duplicating machines. For teachers in these schools some other means of applying the test must be found. Any one of the following methods may be used. First, the entire test may be copied word for word by the pupils and then marked. This is tedious and time-consuming. Second, the entire test may be written on the blackboard by the teacher. Each pupil could number a blank page of paper to correspond to the numbered statements if it is a True-False test, and then write *True* or *False* after the appropriate numbers. The only objection to this suggestion is the inconvenience of writing all the statements on the blackboard. Third, the pupils may be asked to copy on blank paper, 1, 2, 3, and so on, according to the number

of statements. The teacher can then read orally statement 1 and instruct the pupils to make a check after the number 1 on their paper if the statement is true, but to make a cross if the statement is false. This is easily the most convenient way to give the examination. The chief objection to this final method is the difficulty some pupils have in comprehending statements presented orally, particularly if they are long and complicated. When the statement is presented visually the pupil has an opportunity to go back to it enough times to exhaust his possibility of understanding it. By one or another of these methods it is possible for any teacher anywhere to make use of this type of examination. And similarly for tests with three or more choices.

9. Teacher-Made Tests Should Be Designed with Care and Used Year after Year.—Using Book Two as a guide, the teacher should carefully prepare a limited number of excellent items each year. These might well be written on cards—one card per item—so as to facilitate refinement after trial and pupil criticism, elimination, substitution, and regrouping.

10. Tests Should Be Considered from the Point of View of Practicality.—Many of the preceding criteria have suggested certain common-sense conveniences. To these should be added considerations such as cost, the time required from the teacher and pupils, whether the time allowance on the test fits the time allotments for school periods, and whether the test demands more technical competence than is available.

For a special treatment of the construction of tests and examinations in secondary school subjects the reader is referred to:

Hawkes, Lindquist, and Others, *The Construction and Use of Achievement Examinations*, Houghton Mifflin Co., New York, 1936.

Kelley, Turman Lee and Krey, A. C., *Tests and Measurements in the Social Sciences*, C. Scribner's Sons, New York, 1934.

CHAPTER VII

COMPREHENSIVE LIST OF TESTS AND TEST PUBLISHERS

The following comprehensive list of readily obtainable tests presents to the novice, who desires to use a few tests only, a serious problem in selection. The following suggestions will aid him in making a selection, but they are not sufficient to guarantee the *best* selection:

1. *Give the preference to a test by a well-known author.*—It is not enough to look for a well-known author. He should be a well-known *test* author. The very fact that he is well known is a fair guarantee that the test has been constructed by a competent person with enough experience to have some sense of both theoretical and practical considerations.

2. *Give the preference to well-known publishers and distributors of popular tests.*—Such publishers insure competent editing of tests they publish and such distributors select with some discrimination the tests which they distribute.

3. *Send for a descriptive catalog of tests and examine descriptions of possible tests.*—The catalog should give information about cost, number of equivalent forms, grade level to which a test is adapted, *et cetera*.

4. *Send for a sample test and accompanying manual, etc., and apply the criteria developed in Book Two.*

5. *Inquire particularly whether the test is accompanied by a free or inexpensive manual and a table for reading both grade and age scores.*—For preschool children, age scores alone are sufficient, and for college students percentile or sigma scores are sufficient.

6. *Give the preference to recently published tests.*—Even distinguished authors published tests many years ago of which they are not now particularly proud.

7. *Remember that this is a list for test builders as well as for test users.*—Many of these tests should be withdrawn from circulation but not until certain good features in them are built into better tests.

More information may be secured for many of the tests in Table 1 by consulting this comprehensive reference:

Hildreth, Gertrude, *Bibliography of Mental Tests and Rating Scales*, The Psychological Corporation, New York, 1933.

An exhaustive list of annotated character and personality tests, both those readily available and those scattered in publications in the United States and foreign countries, may be found in this reference:

Maller, J. B., *Character and Personality Tests* (Revised), Bureau of Publications, Teachers College, Columbia University, New York, 1937.

Each year's revision of the list presented in Table 1 may be secured for about twenty-five cents from the Guidance Clinic, Teachers College, Columbia University, New York.

Oscar K. Buros, School of Education, Rutgers University, New Brunswick, New Jersey, publishes an annual list and review of new tests and books on tests.

TABLE 1

A BIBLIOGRAPHY OF TESTS, MANUALS, AND SCORING KEYS ON FILE IN AND ASSEMBLED BY THE GUIDANCE LABORATORY AT TEACHERS COLLEGE, COLUMBIA UNIVERSITY

TABLE OF CONTENTS

SUBJECT	PAGE No.	SUBJECT	PAGE No.
Achievement	97	Domestic Science	112
Addresses of Publishers	128	Drawing	116
Agriculture	111		
Algebra	100	Economics	115
Arithmetic	98	Education	123
Art	116	English	
Attitude and Opinion	117	Elementary	104
		High School and College	105
Biology	111	Environment	119
Bookkeeping	127		
Botany	112	Foreign Language	109
		French	109
Chemistry	112		
Civics	115	Geography	115
Clerical Ability	126	Geometry	100
Commercial Law (Voca-		German	110
tional)	125	Government	115
Composition	107		
Culture	96	Handwriting	107
Curriculum	122	Health	118

TABLE 1 (*Continued*)TABLE OF CONTENTS (*Continued*)

SUBJECT	PAGE No.	SUBJECT	PAGE No.
History		Performance	127
American	113	Personality	
English	114	Elementary	119
European	114	High School and College	121
Ancient and Medieval	114	Physics	113
World	115	Publishers, List of	128
Home Economics	112		
		Reading	
Industrial Arts	112	Elementary	101
Information and General		High School and College	103
Culture	96	Religion	119
Intelligence		Reputation	119
Individual	93		
Primary Group	94	Safety	119
Intermediate Group	94	Science	110
Adult Group	95	Sewing	122
Interest Scales	125	Spanish	110
		Spelling	106
Laboratory	124	Stenography	127
Latin	110		
Literature	108	Trigonometry	101
		Typewriting	127
Mathematics	99		
Mechanical Ability	124	Vision	96
Music	116	Vocabulary	106
		Vocational	125
Non-Language	96		
Nursing (Vocational)	125	Zoology	113

*Intelligence Tests**Individual Tests:*

Curtis Point Scale (Stoelting)
 Detroit Tests of Learning Aptitude (Public School Publishing Co.)
 Herring Revision of Binet-Simon Test (World Book)
 Iowa Tests for Young Children (State Univ. of Iowa)
 Kent Emergency Test (E-G-Y) (Psychological Corporation)
 Merrill-Palmer Test (R. Stutsman)
 Minnesota Preschool Scale (Educational Test Bureau)
 Sangren Information Tests for Young Children (World Book)
 Stanford Revision of Binet-Simon Test (Houghton Mifflin)
 Van Alstyne Picture Vocabulary Test for Preschool Children (Public School Publishing Co.)
 Yerkes Point Scale (Stoelting)

Group Tests—Primary Level:

California Test of Mental Maturity for Grades 1-3 (Southern California School Book Depository)
Cleveland Kindergarten Classification Test (Public School Publishing Co.)
Cole-Vincent Group Intelligence Test for School Entrants (Kansas State Teachers College)
Dearborn Intelligence Test I (Educational Test Bureau)
Detroit Beginning First-Grade Intelligence Test (World Book)
Detroit Intelligence Tests (Public School Publishing Co.)
Detroit Kindergarten Test (World Book)
Haggerty Intelligence Examination, Delta I (World Book)
Holley Picture Completion Test for Primary Grades (Public School Publishing Co.)
Indiana Primer Scale (Indiana University)
Institute of Educational Research CAVD Test—Levels A to E (Teachers College Bureau of Publications)
Kingsbury Primary Group Intelligence Test (Public School Publishing Co.)
Kuhlmann-Anderson Intelligence Test (Educational Test Bureau)
Metropolitan Readiness Test (World Book)
Otis Group Intelligence Test, Primary Form (World Book)
Otis Quick Scoring Mental Ability Test (World Book)
Pintner-Cunningham Primary Mental Test (World Book)
Pressey Primary Classification Test (Public School Publishing Co.)
Rhode Island Intelligence Test (Public School Publishing Co.)

*Intelligence Tests**Group Tests—Intermediate Level:*

California Test of Mental Maturity for Grades 4-8 (Southern California School Book Depository)
Dearborn Intelligence Test II (Educational Test Bureau)
Detroit Alpha Intelligence Test (Public School Publishing Co.)
George Washington University International Intelligence Test (Center for Psychological Service)
Goodenough Intelligence Test (World Book)
Haggerty Intelligence Examinations (World Book)
Haines Alertness Test (Stoelting)
Henmon-Nelson Test of Mental Ability for Grades 3-8 (Houghton Mifflin)
Henmon-Nelson Test of Mental Ability for Grades 7-12 (Houghton Mifflin)
Illinois General Intelligence Tests I and II (Public School Publishing Co.)
Kuhlmann-Anderson Intelligence Tests (Educational Test Bureau)
Laycock: Mental Ability Scale (University of Saskatchewan)
McCall Intelligence Test (Laidlaw Bros.)

Otis Self-Administering Test, Higher Form (World Book)
 Pressey Senior Classification Test (Public School Publishing Co.)
 Schubert Information Test (Transient Center)
 Teachers College Psychological Examination (State Teachers College, Minnesota)
 Teachers College Psychological Examination (Illinois State Normal School)
 Thorndike Intelligence Examination for College Entrance (Teachers College Bureau of Publications)
 Thorndike Intelligence Examination for High School Graduates (Teachers College Bureau of Publications)
 Thurstone Psychological Test IV (Stoelting)
 Thurstone Psychological Examination for High School Graduates and College Freshmen (American Council on Education)

Non-Language and Non-Verbal Tests:

Army Beta Test (Psychological Corporation)
 Cattell Group Intelligence Test (Harrap)
 Michigan Non-Verbal Test (University of Michigan)
 Myers Mental Measure (Newson)
 Myers Pantomime Group Intelligence Test (Newson)
 Pintner Non-Language Primary Mental Test (Teachers College Bureau of Publications)
 Pintner Non-Language Mental Test (College Book Co.)
 Porteus Mazes (Stoelting)
 Rockwell: Non-Verbal Perception Scale (Educational Test Bureau)
 Slight Non-Verbal Intelligence Test (Harrap)

Vision Tests

Betts: Telebinocular (Keystone View Co.)
 Jensen: Test for Color Blindness, Visual Acuity, and Astigmatism (Psychological Corporation)
 Ophthalmograph (American Optical Company)

General Information and Background Tests:

Coöperative Contemporary Affairs Test (Educational Records Bureau)
 Coöperative General Culture Test (Educational Records Bureau)
 Detroit Cardinal Objectives Examination (Board of Education, Detroit)
 George Washington Scholastic Aptitude Test (Center for Psychological Service)
 Kansas Every Pupil Test in Contemporary Problems (Kansas State Teachers College)
 Kelty-Moore Test of Concepts in Social Studies (Scribners)
 Kepner Background in Social Studies Test (Ginn)
 Miller Test on French Life and Culture (Kansas State Teachers College)

Teachers College General Information Test (Guidance Laboratory)
Time's Current Affairs Test (Time, Inc.)
Wesley Tests in Social and Political Terms (Scribners)

Educational Achievement Tests

Primary:

Metropolitan Achievement Test (World Book)
Pressey Second Grade Attainment Scale (Public School Publishing Co.)
Pressey Third Grade Attainment Scale (Public School Publishing Co.)
Stanford Achievement Test (World Book)

Intermediate and Above:

Columbia Achievement Test (Columbian Test Service)
Coöperative Test Service Examinations (Educational Records Bureau)
Detroit Tests (Board of Education, Detroit)
Iowa Every Pupil Tests in School Subjects (State Univ. of Iowa)
Iowa High School Content Examination (State Univ. of Iowa)
Iowa Placement Examinations (State Univ. of Iowa)
McCall-Herring Comprehensive Achievement Tests (Laidlaw Bros.)
Metropolitan Achievement Test, Intermediate Form (World Book)
Metropolitan Achievement Test, Advanced Form (World Book)
Modern School Achievement Tests (Teachers College Bureau of Publications)
Myers-Ruch High School Progress Test (World Book)
Ohio Every Pupil Test (Ohio State Univ.)
Ohio State Scholarship Test for Eighth Grade (Ohio State Univ.)
Ohio State General Scholarship Test for High School Seniors (Ohio State Univ.)
Otis-Orleans Standard Graduation Examination (World Book)
Pintner Educational Achievement Test (Teachers College Bureau of Publications)
Progressive Achievement Tests (Southern California School Book Depository)
Public School Attainment Tests (Public School Publishing Co.)
Public School Attainment Scales for High School Entrance (Public School Publishing Co.)
Public School Correlated Attainment Scale for Grades 7-8 (Public School Publishing Co.)
Sones-Harry High School Achievement Test (World Book)
Stanford Achievement Test (World Book)
The Socially Competent Person (Teachers College Bureau of Publications)
Unit Scales of Attainment (Educational Test Bureau)
Unit Scales of Aptitude (Educational Test Bureau)

Arithmetic:

- Brueckner Curriculum Tests in Arithmetic (Winston)
- Buckingham Scale for Problems in Arithmetic (Public School Publishing)
- Buswell-John Fundamental Processes in Arithmetic (Public School Publishing)
- Clapp-Heubner Number Combination Tests (Houghton Mifflin)
- Clapp Number Combination Tests (Houghton Mifflin)
- Clapp-Young Arithmetic Tests (Houghton Mifflin)
- Clark-Otis-Hatton: Instructional Tests in Arithmetic for Beginners (World Book)
- Cleveland Survey Arithmetic Tests (Public School Publishing Co.)
- Compass Diagnostic Tests in Arithmetic (Scott, Foresman)
- Courtis Standard Practice Tests in Arithmetic (World Book)
- DeMay-McCall Rapid Survey Test in Fractions (Teachers College Bureau of Publications)
- DeMay-McCall Standard Test Lessons in Fractions (Teachers College Bureau of Publications)
- Detroit Arithmetic Tests (Board of Education, Detroit)
- Fowlkes-Goff Practice Tests in Arithmetic (Macmillan)
- Green-Studebaker-Knight-Ruch Problem Solving Exercise Cards (Scott, Foresman)
- Hildreth: Arithmetic Achievement Tests (Teachers College Bureau of Publications)
- Hildreth Arithmetic Analysis Tests (G. Hildreth)
- Hotz First Year Algebra Scales (Teachers College Bureau of Publications)
- Institute of Educational Research Arithmetic Problems (Teachers College Bureau of Publications)
- Iowa Every Pupil Test in Basic Arithmetic Skills (State University of Iowa)
- Kinney Scales in Commercial Arithmetic (Public School Publishing)
- Lee Maintenance Drills in Arithmetic (Scott, Foresman)
- Lennes Work, Drill, and Test Sheets in Arithmetic (Laidlaw)
- Los Angeles Diagnostic Arithmetic Tests (Southern California School Book Depository)
- Los Angeles Diagnostic Arithmetic Reasoning Test (Research Service Co.)
- Los Angeles Fundamentals of Arithmetic Test (Research Service Co.)
- Lunceford Diagnostic Test in Addition (Kansas State Teachers College)
- Monroe Diagnostic Tests in Arithmetic (Public School Publishing)
- Monroe Standardized Reasoning Test in Arithmetic (Public School Publishing)
- Monroe Standardized Arithmetic Scales (Public School Publishing)
- New York Survey Tests in Arithmetic (Board of Education, New York City)

-
- Ohio Every Pupil Test in Arithmetic (Ohio State Dept. of Education)
- Otis Arithmetic Reasoning Test (World Book)
- Pett-Dearborn Progress Tests in Arithmetic (Harvard University)
- Pittsburgh Arithmetic Scale (Public School Publishing)
- Plymouth Educational Tests in Arithmetic (Plymouth Press)
- Progressive Arithmetic Tests (Southern California School Book Depository)
- Public School Achievement Test in Arithmetic Reasoning (Public School Publishing)
- Reavis-Breslich Diagnostic Tests in Fundamental Operations of Arithmetic and Problem Solving (University of Chicago)
- Rogers Test for Diagnosing Mathematical Ability (Teachers College Bureau of Publications)
- Sangren-Reidy Instructional Tests in Arithmetic (Public School Publishing)
- Schorling-Clark-Potter Arithmetic Test (World Book)
- Schorling-Sanford Achievement Test in Plane Geometry (Teachers College Bureau of Publications)
- Spencer Diagnostic Arithmetic Tests (C. A. Gregory)
- Staffelbach-Freeland Exercises in Change Making (American Book)
- Stanford Achievement Arithmetic Test (World Book)
- Stevenson Arithmetic Reading Tests (Public School Publishing)
- Stone Reasoning Tests in Arithmetic (Teachers College Bureau of Publications)
- Stone-Hopkins-Brownfield Inventory Tests in Arithmetic (Sanborn)
- Studebaker Practice Exercises in Arithmetic (Scott, Foresman)
- Thompson-Kroner Business Arithmetic Test (Prentice-Hall)
- Upton Arithmetic Workbooks (American Book)
- Upton Inventory Test in Arithmetic (Teachers College Bureau of Publications)
- Wildeman: Test in Common Fractions (Plymouth Press)
- Wilson General Survey Test in Arithmetic (University Publishing Co.)
- Winnetka Speed Practice and Tests in Arithmetic (Winnetka Individual Materials, Inc.)
- Wisconsin Inventory Tests in Arithmetic (Public School Publishing)
- Woody-McCall Mixed Fundamentals (Teachers College Bureau of Publications)
- Woody Arithmetic Scale (Teachers College Bureau of Publications)
- Woody Division Scale B (Teachers College Bureau of Publications)

Mathematics:

- Coöperative General Mathematics Test (Educational Records Bureau)
- Detroit Mathematics Examination (Board of Education, Detroit)
- Iowa Placement Examination in Mathematical Aptitude (State University of Iowa)

Iowa Placement Examination in Mathematical Training (State University of Iowa)
Progressive Mathematics Tests (Southern California School Book Depository)
Rogers Test of Mathematical Ability (Teachers College Bureau of Publications)
Schorling-Reeve Chapter Tests in General Mathematics (Ginn)
Tyler: Mathematics Test (Ohio State Dept. of Education)

Algebra:

Coleman Scale for Testing Ability in Algebra (University of Nebraska)
Columbia Research Bureau Algebra Test (World Book)
Comprehensive Objective Tests in Algebra (Harlow)
Coöperative Algebra Test (Educational Records Bureau)
Coöperative Intermediate Algebra (Educational Records Bureau)
Detroit Algebra Examination (Board of Education, Detroit)
Douglas Standard Diagnostic Tests for Elementary Algebra (University of Cincinnati)
Garman-Schrammel Algebra Test (Kansas State Teachers College)
Goff Algebra Monthly Survey Tests (Palmer Co.)
Hotz First Year Algebra Scales (Teachers College Bureau of Publications)
Illinois Algebra Scales (Public School Publishing)
Institute of Educational Research Elementary Algebra Test (Teachers College Bureau of Publications)
Iowa Every Pupil Test in Algebra (State University of Iowa)
Lee Test of Algebraic Ability (Public School Publishing)
Multiple-Purpose Objective Tests in Algebra (Webb-Duncan)
Nyberg Tests and Drills in First Year Algebra (American Book)
Ohio Every Pupil Test in Algebra (Ohio State Dept. of Education)
Orleans Algebra Prognosis Test (World Book)

Geometry:

American Council Solid Geometry Test (World Book)
Becker-Schrammel Plane Geometry Test (Kansas State Teachers College)
Chandler: Solid Geometry Test (Purdue University)
Columbia Research Bureau Plane Geometry Test (World Book)
Comprehensive Objective Tests in Plane Geometry (Harlow)
Comprehensive Objective Tests in Solid Geometry (Harlow)
Coöperative Geometry (Educational Records Bureau)
Coöperative Plane Geometry Test (Educational Records Bureau)
Coöperative Solid Geometry (Educational Records Bureau)
Detroit Geometry Examination (Board of Education, Detroit)
Greene Plane Geometry Tests (Turner E. Smith)
Iowa Every Pupil Test in Geometry (State University of Iowa)
Iowa Plane Geometry Test (State University of Iowa)

Lane-Knight-Ruch Geometry Rapid Drill Cards (Scott, Foresman)
Lee Geometric Aptitude Test (Southern California School Book
Depository)
McMindes Achievement Test in Plane Geometry (Public School
Publishing)
Multiple-Purpose Objective Test in Geometry (Webb-Duncan)
Minnick Geometry Tests (Public School Publishing)
Ohio Every Pupil Test in Plane Geometry (Ohio State Dept. of
Education)
Orleans Geometry Prognosis Test (World Book)
Orleans Plane Geometry Achievement Test (World Book)
Seattle Geometry Test Series (Public School Publishing)
Totten: Plane Geometry Test (Purdue University)
Webb Geometry Tests (Public School Publishing)
Welte-McKnight Geometry Work Book (Scott, Foresman)

Trigonometry:

American Council Trigonometry Test (World Book)
Coöperative Trigonometry Test (Educational Records Bureau)

*Reading Tests**Elementary School Level:*

Bennett: First Grade Entrance Test in Reading and Intelligence
(Follett Publishing Co.)
Betts Ready to Read Tests (Keystone View Co.)
Burgess Scale for Measuring Ability in Silent Reading (Russell
Sage Foundation)
Chapman Unspeeded Reading Comprehension Test (Lippincott
Co.)
Chapman-Cook Speed of Reading Test (Lippincott Co.)
Clark Reading Readiness Test (Row, Peterson and Co.)
De Garvey Primary Reading Test (Southern California School
Book Depository)
Detroit Reading Tests (World Book)
Detroit Word Recognition Test (World Book)
Dolch-Gray: Basic Reading Tests (Scott, Foresman)
Emporia Silent Reading Test (Kansas State Teachers College)
Garvey Primary Reading Test (Southern California School Book
Depository)
Gates Diagnostic Reading Tests (Teachers College Bureau of Pub-
lications)
Gates Primary Reading Tests (Teachers College Bureau of Publica-
tions)
Gates Silent Reading Tests for Grades 3-8 (Teachers College Bu-
reau of Publications)
Gates Summary of Diagnosis in Reading (Teachers College Bureau
of Publications)

-
- Gates-Ayer: Golden Leaves Work-Play Book (Macmillan Co.)
Gates-Ayer: Magic Hour Work-Play Book (Macmillan Co.)
Gates-Huber: Round the Year Work-Play Book (Macmillan Co.)
Gates-Huber: Peter and Peggy Work-Play Book (Macmillan Co.)
Gates-Huber: Friendly Stories Work-Play Book (Macmillan Co.)
Gates-Pearson Practice Exercises in Reading (Teachers College
Bureau of Publications)
Good Reading Work Cards (Charles Scribner's Sons)
Gray Oral Reading Check Tests (Public School Publishing Co.)
Gray Oral Reading Paragraphs Test (Public School Publishing
Co.)
Greene-Noar Self-Diagnostic Reading Tests (D. C. Heath)
Haggerty Reading Test, Sigma 1 and 3 (World Book)
Hildreth First Grade Reading Test (G. Hildreth)
Hildreth Diagnostic Reading Tests (G. Hildreth)
Hill Test of Word Meanings for Primary Grades (Public School
Publishing Co.)
Ingraham-Clark Reading Tests (Southern California School Book
Depository)
Iowa Silent Reading Test, Elementary Form (World Book)
Iowa Every Pupil Test in Silent Reading Comprehension (State
Univ. of Iowa)
Kansas Silent Reading Tests I and II (Bureau of Educational
Measurements and Standards, State Normal School, Emporia,
Kansas)
Lee-Clark Primer Reading Test (Southern California School Book
Depository)
Lee-Clark First Reader Test (Southern California School Book
Depository)
Lee-Clark Reading Readiness Test (Southern California School
Book Depository)
Los Angeles Elementary Reading Test (Southern California School
Book Depository)
McCall-Crabbs: Standard Test Lessons in Reading (Teachers Col-
lege Bureau of Publications)
McGaughy Informal Reading Tests (Ginn)
Manmiller: Word Recognition Test (World Book)
Michigan Speed of Reading Test (Psychological Corporation)
Monroe, M., Diagnostic Reading Examination (Stoelting)
Monroe Reading Aptitude Test (Houghton Mifflin Co.)
Monroe Standardized Silent Reading Tests I and II (Public School
Publishing Co.)
Nelson Silent Reading Test (Houghton Mifflin)
O'Rourke Survey Tests of Reading (Psychological Corporation)
Philadelphia Reading Tests (Board of Education, Philadelphia)
Phillips-Woody Group Test for Reversals (University of Michigan)
Pressey First Grade Word Reading Tests (Public School Publish-
ing Co.)

Pressey Diagnostic Tests in Reading (Public School Publishing Co.)
Pressey Reading Selections: Mechanics of Reading (Public School Publishing Co.)
Pressey-Grant First Grade Reading Test (Public School Publishing Co.)
Price Practical Oral Reading Test (E. D. Price)
Progressive Reading Tests (Southern California School Book Depository)
Purdue Reading Test (Lafayette Printing Co.)
Sangren-Woody Reading Tests (World Book)
Sangren-Wilson Instructional Tests in Reading (Public School Publishing Co.)
Shank Tests of Reading Comprehension I and II (C. A. Gregory)
Stone: Classification Test for Beginners in Reading (Webster Publishing Co.)
Stone Narrative Reading Tests (Public School Publishing Co.)
Study Type of Reading Exercises (Teachers College Bureau of Publications)
Survey Reading Tests (Teachers College Bureau of Publications)
Thorndike-McCall Reading Scales (Teachers College Bureau of Publications)
Traxler Silent Reading Test (Public School Publishing Co.)
Van Wagenen Reading Readiness Test (Educational Test Bureau)
William: Primary Reading Test (Public School Publishing Co.)
William Reading Test for Grades 4-9 (Public School Publishing Co.)

High School and College Level:

Buffalo Reading Test (University of Buffalo)
Chapman Unspeeded Reading Comprehension Test (Lippincott Co.)
Haggerty Reading Test, Sigma 3 (World Book)
Institute of Educational Research Speed of Reading Test (Teachers College Bureau of Publications)
Iowa Comprehension Tests (State University of Iowa)
Iowa Silent Reading Test, Revised Form (World Book)
Kansas Silent Reading Test III (State Normal School, Emporia, Kansas)
Knode XYZ College Freshman Reading Test (University of New Mexico)
Michigan Speed of Reading Tests (Psychological Corporation)
Minnesota Speed of Reading Tests for College Students (University of Minnesota)
Minnesota Reading Examination for College Students (University of Minnesota)
Monroe Standardized Silent Reading Tests, .III (Public School Publishing Co.)
Mount Holyoke Reading Test (Mt. Holyoke College)
Nelson-Denny Reading Tests (Houghton Mifflin)
Ohio State Reading Comprehension Test (Ohio State University)

Ohio State Study Performance Test (Ohio State University)
O'Rourke Survey of Reading (Psychological Corporation)
Pressey General Reading Test (Ohio State Dept. of Education)
Pressey Special Reading Test (Ohio State Dept. of Education)
Pressey Test on Reading Comprehension (Ohio State Dept. of Education)
Progressive Reading Test (Southern California School Book Depository)
Purdue Reading Test (Lafayette Printing Co.)
Shank Tests of Reading Comprehension III (C. A. Gregory)
Thorndike-McCall Reading Scale (Teachers College Bureau of Publications)
Traxler Silent Reading Test (Public School Publishing Co.)
Whipple High School and College Reading Tests (Public School Publishing Co.)
Wrenn: Study Habits Inventory (Stanford University Press)
Wrenn: Practical Study Aids (Stanford University Press)

English Tests

Elementary School Level:

Bregman Language Completion Scale (Psychological Corporation)
Briggs English Forms (Teachers College Bureau of Publications)
Charters Diagnostic Language Test (Public School Publishing Co.)
Charters Diagnostic Language and Grammar Test (Public School Publishing Co.)
Clapp Test for Correct English (Houghton Mifflin)
Clapp-Young English Test (Houghton Mifflin)
Clark Letter Writing Test (Public School Publishing Co.)
Cleveland English Composition and Grammar Test (Houghton Mifflin)
Detroit English Examination (Board of Education, Detroit)
Detroit English Test—Grammatical Forms (Board of Education, Detroit)
Franseen Diagnostic Tests in Language (C. A. Gregory)
Iowa Elementary Language Test (Educational Test Bureau)
Iowa Grammar Information Test (State University of Iowa)
Iowa Every Pupil Test in Basic Language Skills (State University of Iowa)
Los Angeles Diagnostic Language Test (Southern California School Book Depository)
New York English Survey Test in Language Usage (Public School Publishing Co.)
New York English Survey Test in Grammar (Public School Publishing Co.)
New York English Test in Sentence Structure (Public School Publishing Co.)

Ohio Every Pupil Test in English (Ohio State Dept. of Education)
O'Rourke Achievement Test in English (Educational and Personnel
Publishing Co.)
Philadelphia Test in Outlining (Board of Education, Philadelphia)
Plymouth Educational Test in Punctuation (Plymouth Press)
Pribble-McCrory Diagnostic Elementary Language Test (Lyons &
Carnahan)
Pribble-McCrory Diagnostic Tests in Practical English Grammar
(Lyons and Carnahan)
Progressive Language Tests (Southern California School Book
Depository)
Purdue Diagnostic English Test (Lafayette Printing Co.)
Stanford Language Usage Test (World Book Co.)
Starch Punctuation Scale (Public School Publishing Co.)
Trabue-Kelley Language Completion Exercises (Teachers College
Bureau of Publications)
Wilson Language Error Test (World Book Co.)

High School and College Level:

Barrett-Ryan-Schrammel English Test (World Book)
Barrett-Ryan English Test (Kansas State Teachers College)
Charters Diagnostic Language Test (Public School Publishing Co.)
Charters Diagnostic Language and Grammar Test (Public School
Publishing Co.)
Clapp Test for Correct English (Houghton Mifflin)
Clapp-Young English Test (Houghton Mifflin)
Clark Letter Writing Test (Public School Publishing Co.)
Clatworthy Library Test for College Students (L. M. Clatworthy)
Cleveland English Composition and Grammar Test (Houghton
Mifflin Co.)
Columbia Research Bureau English Test (World Book)
Comprehensive Objective Tests in Grammar and Composition
(Harlow)
Comprehensive Objective Tests in Correct English Usage (Harlow)
Coöperative English Test (Educational Records Bureau)
Cross English Test (World Book)
Davis Tests in English Fundamentals (Ginn)
Exercises in the Appreciation of Poetry (Teachers College Bureau of
Publications)
George Washington University Language Aptitude Test (Center
for Psychological Service)
Iowa English Organization Test (State Univ. of Iowa)
Iowa Every Pupil Test in English Correctness (State Univ. of Iowa)
Iowa Grammar Information Test (State Univ. of Iowa)
Iowa Placement Examination in English Training (State Univ. of
Iowa)
Iowa Placement Examination in English Aptitude (State Univ. of
Iowa)

-
- McClusky-Dolch Study Outline Test (Public School Publishing Co.)
 Multiple Purpose Tests in Grammar and Composition (Webb-Duncan)
 Nassau County Supplement to the Hillegas Composition Scale (Teachers College Bureau of Publications)
 Nelson High School English Test (Houghton Mifflin)
 Ohio Every Pupil Test in English (Ohio State Dept. of Education)
 O'Rourke Achievement Test in English Usage (Educational and Personnel Publishing Co.)
 Poley: Precis Test (Public School Publishing Co.)
 Pressey English Survey Test (Ohio State Dept. of Education)
 Pribble-McCrory Diagnostic Tests in Practical English Grammar (Lyons and Carnahan)
 Progressive Language Tests for H. S. and College Students (Southern California School Book Depository)
 Purdue English Test (Lafayette Printing Co.)
 Purdue Diagnostic English Test (Lafayette Printing Co.)
 Purdue Placement Test in English (Houghton Mifflin)
 Rinsland-Beck Natural Test for English Usage (Public School Publishing Co.)
 Scott-Reed-Weideman English Classification Test (Kansas State Teachers College)
 Schutte Diction Test (Public School Publishing Co.)
 Shepherd English Test (Houghton Mifflin Co.)
 Starch Punctuation Scale (Public School Publishing Co.)
 Steeves Placement Test in English for College Students (Columbia University)
 Trabue-Kelley Language Completion Exercises (Teachers College Bureau of Publications)
 Tressler: English Minimum Essentials Test (Public School Publishing Co.)
 Wakefield Diagnostic English Test (C. A. Gregory)
 Welling English Composition and Grammar Test (Houghton Mifflin)
 Wilson Language Error Test (World Book)

Spelling and Vocabulary Scales

- Armstrong-Danielson: Sentence Vocabulary Test (Southern California School Book Depository)
 Ayres Spelling Scale (Russell Sage Foundation)
 Buckingham Extension of the Ayres Scale (Public School Publishing Co.)
 Davis-Schrammel: Spelling Test (Kansas State Teachers College)
 Detroit Vocabulary Test (Board of Education, Detroit)
 Gates-Russell Spelling Diagnosis Test (Teachers College Bureau of Publications)
 Guy Spelling Scale (Public School Publishing Co.)

Holley Sentence Vocabulary Test (Public School Publishing Co.)
Inglis Tests of English Vocabulary (Ginn)
Iowa Every Pupil Test in Vocabulary and Basic Study Skills (State University of Iowa)
Iowa Spelling Scales (State University of Iowa)
Kansas Every Pupil Test in Spelling (Kansas State Teachers College)
Kennon Test of Literary Vocabulary (Teachers College Bureau of Publications)
Markham English Vocabulary Test (Public School Publishing Co.)
Minnesota College Aptitude Test (University of Minnesota Press)
Monroe Timed Spelling Test (Public School Publishing Co.)
Morrison-McCall Spelling Scale (World Book)
New York Spelling Tests (Board of Education, New York City)
O'Rourke Survey Test of Vocabulary (Psychological Institute)
Philadelphia Index and Dictionary Test (Board of Education, Detroit)
Plymouth Educational Tests in Vocabulary (Plymouth Press)
Pressey Test of Technical Vocabularies (Public School Publishing Co.)
Thorndike Test of Word Knowledge (Teachers College Bureau of Publications)
Turner-Miller: Cross Word Puzzle Speller (Public School Publishing Co.)
Van Wagenen Unit Scales of Attainment in Spelling (Educational Test Bureau)

Composition Scales

Cross Diagnostic English Composition Test (Little, Brown)
Detroit Composition Examination (Board of Education, Detroit)
Driggs-Mayhew National Scales for Measuring Composition (University Publishing Co.)
Hillegas Scale for Measuring English Composition (Teachers College Bureau of Publications)
Hudelson Typical Composition Ability Scale (Public School Publishing Co.)
Lewis English Composition Scales (World Book)
Pressey-Conklin: Student's Guide to Correctness in Written Work (Public School Publishing Co.)
Pressey-Bowers Diagnostic Tests in English Composition (Public School Publishing Co.)
Seaton-Pressey Minimum Essential Test in English Composition (Public School Publishing Co.)
Van Wagenen English Composition Scales (World Book)

Handwriting Scales

Ayres Handwriting Scale (Russell Sage Foundation)
Cleveland Business Penmanship Scale (Harter School Supply Co.)
Conrad Manuscript Writing Standards (Teachers College Bureau of Publications)

Courtis Standard Practice Tests in Handwriting (World Book)
 Detroit Handwriting Test (Board of Education, Detroit)
 Freeman Chart for Diagnosing Faults in Handwriting (Houghton Mifflin)
 Freeman Handwriting Measuring Scale for Grades 7, 8 and 9 (Zaner-Bloser)
 Leamer Practice Sentences in Handwriting (Public School Publishing Co.)
 Metropolitan Primary Cursive Handwriting Scale (Teachers College Bureau of Publications)
 Newland Chart for Diagnosis of Illegibilities in Written Arabic Numerals (Public School Publishing Co.)
 Thorndike Scale for Handwriting of Children (Teachers College Bureau of Publications)
 West Handwriting Scale (Palmer Co.)

Literature and Appreciation of Literature Tests

Abbott-Trabue Exercises in Judging Poetry (Teachers College Bureau of Publications)
 Analytical Scales of Attainment in Literature (Educational Test Bureau)
 Carroll Prose Appreciation Test (Educational Test Bureau)
 Cavins Test in Poetry (Public School Publishing Co.)
 Comprehensive Objective Tests in American Literature (Harlow)
 Comprehensive Objective Tests in English Literature (Harlow)
 Coöperative Literary Acquaintance Test (Educational Records Bureau)
 Coöperative Literary Comprehension Test (Educational Records Bureau)
 Detroit Literature Appreciation Test (Board of Education, Detroit)
 Gehlman American Literature Test (Harcourt, Brace)
 George Washington University English Literature Tests (Center for Psychological Service)
 Hahn: Tests on English Classics (Houghton Mifflin)
 Inglis English Literature Tests (Harcourt, Brace)
 Iowa Every Pupil Test in American Literature (State University of Iowa)
 Iowa Every Pupil Test in English Literature (State University of Iowa)
 Lagosa-Wright Literature Appreciation Test (Public School Publishing Co.)
 Logan-Parks Literary Background Test (Heath)
 Multiple-Purpose Objective Tests in American Literature (Webb-Duncan)
 Multiple-Purpose Objective Tests in History of English Literature (Webb-Duncan)
 Multiple-Purpose Objective Tests in the Classics (Webb-Duncan)
 New York English Survey Test in Literature Information (Public School Publishing Co.)

Odell Scale for Rating Pupils' Answers to Nine Types of Thought Questions in English Literature (University of Illinois)
Ohio Every Pupil Test in American and English Literature (Ohio State Dept. of Education)
Plymouth English Tests in Literature (Plymouth Press)
Stanford American Literature Test (C. A. Gregory)
Stanford English Literature Test (C. A. Gregory)
Stanford Tests in Comprehension of Literature (Stanford University Press)
Wykoff: Understanding and Appreciation of Poetry (Purdue University)

Foreign Language Tests

General Tests in Foreign Language:

Handschin Language Predetermination Test (C. H. Handschin)
Hoffman Bilingual Schedule (Teachers College Bureau of Publications)
Iowa Placement Examination in Foreign Language Aptitudes (State University of Iowa)
Luria-Orleans Modern Language Prognosis Test (World Book)
Symonds Foreign Language Prognosis Test (Teachers College Bureau of Publications)
Wilkins Prognosis Test in Modern Language (World Book)

French:

American Council Alpha French Test (World Book)
American Council Beta French Test (World Book)
American Council French Grammar Test (World Book)
Broom-Brown: Silent Reading Test in French (Southern California School Book Depository)
Columbia Research Bureau French Test (World Book)
Comprehensive Objective Tests in French (Harlow)
Coöperative French Test (Educational Records Bureau)
Detroit French Examination (Board of Education, Detroit)
Fowlkes-Young: Instructional Tests in French (Houghton Mifflin)
Handschin Modern Language Test in French (World Book)
Harvard French Vocabulary Test (Ginn)
Henmon French Tests (World Book)
Iowa Placement Examination in French Training (State University of Iowa)
Lundeberg-Thorp: Audition Test in French (Ohio State University)
Miller-Davis: French Test (Kansas State Teachers College)
Multiple-Purpose Objective Tests in French (Webb-Duncan)
Ohio Every Pupil Test in French (Ohio State University)
Sammartino-Krause Standard French Test (Public School Publishing Co.)

German:

Aiken-Held First Year German Test (Kansas State Teachers College)
American Council on Education German Reading Scale (Public School Publishing Co.)
American Council Alpha German Test (World Book)
Columbia Research Bureau German Test (World Book)
Coöperative German Test (Educational Records Bureau)

Latin:

Alexander: First Year Latin Test (Purdue University)
Comprehensive Objective Tests in Latin (Harlow)
Coöperative Latin Test (Educational Records Bureau)
Detroit Latin Examination (Board of Education, Detroit)
Godsey Latin Composition Test (World Book)
Henmon Latin Tests (World Book)
Holtz Latin Tests (Kansas State Teachers College)
Hutchinson Latin Grammar Scale (Public School Publishing Co.)
Lee: Second Year Latin Test (Purdue University)
Lohr-Latshaw Latin Form Test (University of North Carolina)
Multiple-Purpose Objective Tests in Latin (Webb-Duncan)
New York Latin Achievement Test (World Book)
Ohio Every Pupil Test in Latin (Ohio State Dept. of Education)
Orleans-Solomon Latin Prognosis Test (World Book)
Powers Diagnostic Latin Test (Public School Publishing Co.)
Pressey Test in Latin Syntax (Public School Publishing Co.)
Stevenson-Coxe Latin Derivative Test (Public School Publishing Co.)
Stevenson Latin Vocabulary Test (Public School Publishing Co.)
Tyler-Pressey Test in Latin Verbs (Public School Publishing Co.)
White Latin Test (World Book)

Spanish:

American Council Spanish Test (World Book)
Broom Spanish Test (Public School Publishing Co.)
Columbia Research Bureau Spanish Test (World Book)
Comprehensive Objective Tests in Spanish (Harlow)
Coöperative Spanish Test (Educational Records Bureau)
Handschin Modern Language Test in Spanish (World Book)
Iowa Placement Examination in Spanish Training (State Univ. of Iowa)
Multiple-Purpose Objective Tests in Spanish (Webb-Duncan)
Stanford Spanish Test (Stanford University Press)
Wilkins Achievement Test in Spanish (Holt)

*Science Tests**Miscellaneous Science Tests:*

Comprehensive Objective Tests in General Science (Harlow)
Cooperative General Science Test (Educational Records Bureau)

Detroit Elementary Science Examination (Board of Education, Detroit)
Detroit Social Science Test (Board of Education, Detroit)
Downing: Range of Information Test in Science (University of Chicago Press)
Dvorak: General Science Scale (Public School Publishing Co.)
Giles-Thomas-Schmidt: General Science Examinations (State Department of Public Instruction, Madison, Wisconsin)
Iowa Every Pupil Test in General Science (State University of Iowa)
Melbo Social Science Test (World Book)
Michigan Botany Test (Public School Publishing Co.)
Multiple-Purpose Objective Test in General Science (Webb-Duncan)
Odell Scales for Rating Pupils' Answers to Nine Types of Thought Questions in General Science (University of Illinois)
Ohio Every Pupil Test in General Science (Ohio State Dept. of Education)
Powers General Science Test (Teachers College Bureau of Publications)
Public School Achievement Test in Nature Study (Public School Publishing Co.)
Ruch-Popenoe General Science Tests (World Book)
Stanford Scientific Aptitude Test (Stanford University Press)
Van Wagenen Reading Scales in General Science (Public School Publishing Co.)

Agriculture:

Auburn Test for Agricultural Information (Alabama Polytechnic Inst.)
Dickinson Test on Dairy Husbandry Information (Univ. of Missouri)
National Agricultural Tests (Van Cleve Publishers)

Biology:

Catholic High School Tests in Biology (Catholic Education Press)
Comprehensive Objective Tests in Biology (Harlow)
Coöperative Biology Test (Educational Records Bureau)
Coöprider: Information Exercises in Biology (Public School Publishing Co.)
Davis: Biology Tests (Metzer-Bush)
Detroit Biology Examination (Board of Education, Detroit)
Downing-McAtee Biology Unit Tests (Lyons and Carnahan)
Hunter-Kitch Mastery Tests in Biology (American Book Co.)
Multiple-Purpose Objective Tests in Biology (Webb-Duncan)
Oakes and Powers: General Biology Test (Teachers College Bureau of Publications)
Objective Unit Tests on Everyday Problems in Biology (Scott, Foresman)

Ohio Every Pupil Test in Biology (Ohio State Dept. of Education)
Pieper-Beauchamp-Frank: Tests on Everyday Problems in Biology
(Scott, Foresman)
Pressey Biology Test (World Book)
Rifenburgh: Biology Test (Purdue University)
Ruch-Cossman Biology Test (World Book)
Van Wagenen Reading Scales in Biology (Educational Test Bureau)
Williams Biology Test (Kansas State Teachers College)

Botany:

Coöperative Botany Test (Educational Records Bureau)

Chemistry:

Cohn Briscoe: Chemistry Tests (Mentzer-Bush)
Columbia Research Bureau Chemistry Test (World Book)
Comprehensive Objective Tests in Chemistry (Harlow)
Coöperative Chemistry Test (Educational Records Bureau)
Detroit Chemistry Examination (Board of Education, Detroit)
Dull: Chemistry Tests (Holt)
George Washington University General Chemistry Test (Center for Psychological Service)
Glenn-Welton: Instructional Tests in Chemistry (World Book)
Iowa Placement Examination in Chemistry Aptitude (State University of Iowa)
Iowa Placement Examination in Chemistry Training (State University of Iowa)
Multiple-Purpose Objective Test in Chemistry (Webb-Duncan)
Ohio Every Pupil Test in Chemistry (Ohio State Dept. of Education)
Pershing Laboratory Chemistry Test (Public School Publishing Co.)
Powers General Chemistry Test (World Book)
Rich Chemistry Test (Public School Publishing Co.)
Rivett Chemistry Tests (Northwestern High School, Detroit, Michigan)

Domestic Science, Home Economics, and Industrial Arts Tests:

Detroit Domestic Science Test (Board of Education, Detroit)
Detroit Household Mechanics Test (Board of Education, Detroit)
Engle-Stenquist Home Economics Test (World Book)
Frear-Cox Clothing Test (Public School Publishing Co.)
Home Economics Test for Girls Completing 8th Grade (Teachers College Bureau of Publications)
Illinois Information Test on Foods (Public School Publishing Co.)
Leary and Dry: Technical Information Test for Girls (Stoelting)
Multiple-Purpose Objective Tests in Home Economics (Webb-Duncan)
Murdoch Sewing Scale (Teachers College Bureau of Publications)
Murdoch Analytic Sewing Scale for Measuring Separate Stitches (Teachers College Bureau of Publications)

Nash-Van Duzee Industrial Arts Test (Bruce Publishing Co.)
Patrick Industrial Arts Test (Independent Press)
Stevenson-Trilling Tests in Home Economics (Webb-Duncan)
Unit Scales of Attainment in Foods and Household Management
(Educational Test Bureau)
Wells-Lauback: Industrial Arts Test (Manual Arts Press)

Physics:

Columbia Research Bureau Physics Test (World Book)
Comprehensive Objective Tests in H. S. Physics (Harlow)
Coöperative Physics Test for H. S. Students (Educational Records Bureau)
Coöperative Physics Test in Light for College Students (Educational Records Bureau)
Coöperative Physics Test in Mechanics for College Students (Educational Records Bureau)
Coöperative Physics Test in Sound for College Students (Educational Records Bureau)
Coöperative Physics Test in Electricity for College Students (Educational Records Bureau)
Detroit Physics Examination (Board of Education, Detroit)
Fulner-Schrammel Physics Test (Kansas State Teachers College)
Hughes: Physics Scale (Public School Publishing Co.)
Hurd Test in High School Physics (Teachers College Bureau of Publications)
Iowa Achievement Examination in College Physics (State University of Iowa)
Iowa Placement Examination in Physics Aptitude (State University of Iowa)
Iowa Placement Examination in Physics Training (State University of Iowa)
Kilzer-Kirby: Physics Test (Public School Publishing Co.)
Multiple-Purpose Objective Test in Physics (Webb-Duncan)
Ohio Every Pupil Test in Physics (Ohio State Dept. of Education)

Zoölogy:

Coöperative Zoölogy Test (Educational Records Bureau)

History

American:

Barr-Dagett Information Test in American History (Educational Test Bureau)
Carman-Barrows-Wood Junior American History Test (World Book)
Clark Exercises in the Use of Historical Evidence (Scribners)
Columbia Research Bureau American History Test (World Book)
Comprehensive Objective Tests in American History (Harlow)

Coöperative American History Test (Educational Records Bureau)
Dawold: American History Test (Purdue University)
Denny-Nelson American History Test (World Book)
Detroit History Examination (Board of Education, Detroit)
Ely-King Tests in American History (Southern California School Book Depository)
Ely-King Interpretation Tests in American History (Southern California School Book Depository)
Farley Test of Factual Relations in American History (Farley)
Harlan Information Tests in American History (Public School Publishing Co.)
Iowa Every Pupil Test in U. S. History (State University of Iowa)
Iowa General Information Test in American History (Webb-Duncan)
Odell Scales for Rating Pupils' Answers to Nine Types of Thought Questions in American History (University of Illinois)
Ohio Every Pupil Test in American History (Ohio State Dept. of Education)
Patterson Tests on the Federal Constitution (Palmer Co.)
Plymouth Educational Tests in History (Plymouth Press)
Pressey-Richards Tests in American History (Public School Publishing Co.)
Soth-Vannest Proficiency Tests in United States History (Webster Publishing Co.)
Van Wagenen American History Scales (Teachers College Bureau of Publications)
Multiple-Purpose Objective Tests in American History (Webb-Duncan)

European:

American Council European History Test (World Book)
Comprehensive Objective Tests in Modern European History (Harlow)
Coöperative Modern European History Test (Educational Records Bureau)
George Washington University Modern European History Test (Center for Psychological Service)
Multiple-Purpose Objective Tests in Modern European History (Webb-Duncan)
Vannest: Diagnostic Test in Modern European History (Indiana Univ. Bookstore)

English:

Comprehensive Objective Tests in English History (Harlow)
Coöperative English History Test (Educational Records Bureau)

Ancient and Medieval:

Comprehensive Objective Test in Ancient and Medieval History (Harlow)

Coöperative Ancient History Test (Educational Records Bureau)
Coöperative Medieval History Test (Educational Records Bureau)
Multiple-Purpose Objective Tests in Ancient and Medieval History
(Webb-Duncan)

World:

Comprehensive Objective Tests in World History (Harlow)
Coöperative World History Test (Educational Records Bureau)
Dawold World History Test (Purdue University)
Iowa Every Pupil Test in World History (State University of Iowa)
Multiple-Purpose Objective Test in World History (Webb-Duncan)
Ohio Every Pupil Test in World History (Ohio State Dept. of
Education)

Civics and Government Tests

Almack Test of American Civics and Government (Gregory)
American Council Civics and Government Test (World Book)
Brown-Woody Civics Test (World Book)
Burton Civics Test (World Book)
Comprehensive Objective Tests in American Government and Civics
(Harlow)
Comprehensive Objective Tests in Community Civics (Harlow)
Comprehensive Objective Tests in Democracy (Harlow)
Haley: American Government and Civics Test (Harlow)
Hill Test in Civic Attitudes (Public School Publishing Co.)
Hill Test in Civic Information (Public School Publishing Co.)
Hill-Wilson Civic Action Test (Public School Publishing Co.)
Iowa Every Pupil Test in American Government (State University
of Iowa)
Kefauver-Hand Guidance Tests and Inventories (World Book)
Magruder-Chamber-Clinton: American Civics and Government Test
for High Schools (Public School Publishing Co.)
Malan: Junior and Senior H. S. Civics Test (Purdue University)
Mordy-Schrammel: Elementary Civics Test (Kansas State Teachers
College)
Odell Scales for Rating Pupils' Answers to Nine Types of Thought
Questions in Civics (University of Illinois)
Teeter Objective Tests in American Democracy (McGraw-Hill)

Economics Tests

American Council Economics Test (World Book)
Comprehensive Objective Tests in Economics (Harlow)
Iowa Every Pupil Test in Economics (State University of Iowa)

Geography Tests

Branom: Diagnostic Tests in Geography (McKnight and McKnight)
Buckingham-Stevenson Geography Tests on the United States (Public
School Publishing Co.)

Comprehensive Objective Tests in Physical, Industrial, and Commercial Geography (Harlow Publishing Co.)
Hahn-Lackey Geography Scale (Wayne University)
Hill Tests in Physical Geography (Webster Publishing Co.)
Philadelphia Map Reading Test (Board of Education, Philadelphia)
Multiple-Purpose Objective Test in Physical Geography (Webb-Duncan)
Multiple-Purpose Objective Test in Industrial Geography (Webb-Duncan)
Ohio Every Pupil Test in Geography (Ohio State Dept. of Education)
Plymouth Educational Tests in Geography (Plymouth Press)
Posey-Van Wagenen Geography Test (Public School Publishing Co.)
Stevenson-Ridgley-Shipman Geography Test on Asia (Public School Publishing Co.)
Stevenson-Ridgley-Shipman Geography Test on Europe (Public School Publishing Co.)
Stevenson-Ridgley-Shipman Geography Test on South America (Public School Publishing Co.)
Wiederfeld-Walther: Geography Test (World Book Co.)

Art Tests

Detroit Art Test (Board of Education, Detroit)
Detroit Lettering Test (Board of Education, Detroit)
Kline-Carey Measuring Scale in Drawing (Johns Hopkins Press)
Knauber Art Ability Test (University of Cincinnati)
Landis Achievement Test in Printing (R. H. Landis)
Lewerenz Tests in Fundamental Abilities of Visual Art (Southern California School Book Depository)
McAdory Art Test (Teachers College Bureau of Publications)
Meier-Seashore Art Judgment Test (State University of Iowa)
Minnesota House Design and House Furnishing Test (University of Minnesota Press)
Thorndike Drawing Scale (Teachers College Bureau of Publications)
Wells Printing Test (Manual Arts Press)

Music Tests

Beach Music Tests (Kansas State Teachers College)
Bowen Graded Melodies for Individual Sight Singing (Laidlaw)
Cleveland Music Test (Bureau Educational Research, Cleveland)
Courtis Music Test (S. A. Courtis)
Drake Musical Memory Test (Public School Publishing Co.)
Fullerton Standardization Tests in Music for Rural Schools (Follett)
Gildersleeve Music Achievement Test (Teachers College Bureau of Publications)
Hildbrand Sight Singing Test (World Book)
Hutchinson Music Tests (Public School Publishing Co.)
Kelsey Standardized Tests of Musical Accomplishment (C. A. Gregory)
Knuth Achievement Tests in Music (Educational Records Bureau)

-
- Kwalwasser-Dykema Music Tests (Fischer Co.)
 Kwalwasser-Ruch Test of Musical Accomplishment (State University of Iowa)
 Kwalwasser Test of Music Information and Appreciation (State University of Iowa)
 McCauley Examination in Public School Music (Joseph E. Avent)
 Moon Diagnostic Tests in Harmony (Jones)
 Oregon Musical Discrimination Tests (C. H. Stoelting Co.)
 Otterstein-Mosher Sight Singing Test (Stanford University Press)
 Plymouth Educational Test in Music (Plymouth Press)
 Providence Inventory Test in Music (World Book)
 Seashore Music Talent Test (State University of Iowa)
 Torgerson-Fahnestock Music Test (Public School Publishing Co.)
 (For an annotated bibliography of the foregoing and other available tests see, *A Descriptive Bibliography of Prognostic and Achievement Tests in Music* (Teachers College Bureau of Publications))

Attitude and Opinion Scales

- Attitudes S-A Test (Association Press)
 Bruner-Linden: Tentative Check List for Determining the Positions Held by Students on Forty Crucial World Problems (Teachers College Bureau of Publications)
 Cottrell Test on Controversial Issues in Higher Education (Cottrell)
 Critical Thinking in the Social Studies (Teachers College Bureau of Publications)
 Harper, H. R.: Study of Opinions, Feelings, and Attitudes Concerning International Problems (Association Press)
 Harper, M.: A Social Study (Teachers College Bureau of Publications)
 Hart: A Test of Social Attitudes and Interests (State Univ. of Iowa)
 Lentz: C-R Opinionnaire (Character Research Institute)
 Maller-Tuttle: Social Orientation Test (Maller)
 Miller: A Scale for Measuring Attitude toward Any Vocation (Purdue Research Foundation)
 Miller: A Scale for Measuring Attitude toward Teaching (Purdue Research Foundation)
 Minnesota Scale for the Survey of Opinions (University of Minnesota Press)
 Neumann-Kulp-Davidson: Test of International Attitudes (Teachers College Bureau of Publications)
 Noll: What Do You Think? (Teachers College Bureau of Publications)
 Opinions on Race Relations (Association Press)
 Opinions on International Questions (Association Press)
 Palmer: What Do You Think about Orientals in the United States? (Friendship Press)
 Pintner General Opinion Test (mimeographed) (Pintner)

-
- Raup: Teacher's Views on Some Problems in General Educational Theory (Teachers College Bureau of Publications)
- Sweet: Measurement of Personal Attitudes in Younger Boys (Association Press)
- Test of Liberal Thought (Teachers College Bureau of Publications)
- Thurstone: Measurement of Attitude toward God (Univ. of Chicago Press)
- Thurstone: Measurement of Attitude toward the Church (Univ. of Chicago Press)
- Thurstone: Measurement of Attitude toward War (Univ. of Chicago Press)
- Thurstone: Measurement of Attitude toward the Negro (Univ. of Chicago Press)
- Thurstone: Measurement of Attitude toward Birth Control (Univ. of Chicago Press)
- Thurstone: Measurement of Attitude toward Patriotism (Univ. of Chicago Press)
- Thurstone: Measurement of Attitude toward the Bible (Univ. of Chicago Press)
- Thurstone: Measurement of Attitude toward the Germans (Univ. of Chicago Press)
- Watson Test of Public Opinion (Teachers College Bureau of Publications)

Health Attitudes and Information Tests

- American Child Health Association Tests (American Child Health Assoc.)
- Brewer-Schrammel: Health Knowledge and Attitude Test (Kansas State Teachers College)
- C. E. I. Health Knowledge Test (Association Press)
- Franzen-Derryberry-McCall: Health Awareness Test (Teachers College Bureau of Publications)
- Gates-Strang Health Knowledge Test (Teachers College Bureau of Publications)
- I Am Growing Up (Teachers College Bureau of Publications)
- Kefauver-Hand Health Guidance Test (World Book)
- Ohio Every Pupil Test in Health Education and Hygiene (Ohio State Dept. of Education)
- Payne: Habits and Practices in Health and Accident Prevention (Public School Publishing Co.)
- Personal Health Standard and Scale (Teachers College Bureau of Publications)
- Public School Achievement Health Test (Public School Publishing Co.)
- Schrammel-Brewer: Health Knowledge and Attitude Test (Kansas State Teachers College)
- White House Conference Health Blanks (Century Co.)
- Wood-Rowell Health and Growth Record (Teachers College Bureau of Publications)

Safety Tests

Highway Safety Tests (Travelers Insurance Co.)
Pancok: Safety Test for Primary Grades (National Safety Council)
National Safe Drivers Test (National Bureau of Casualty and Safety Underwriters)
National Tests in Safety Education (National Bureau of Casualty and Safety Underwriters)
Perkins: Silent Reading Test on Safety (Travelers Insurance Co.)
Stack: Home Safety Test (Hartford Accident and Indemnity Co.)
What's Wrong with These Drivers and Pedestrians? (Travelers Insurance Co.)

Religious Attitudes and Information Tests

Case Test of Liberal Thought (Teachers College Bureau of Publications)
Laycock Test of Biblical Information (Association Press)
Test of Religious Thinking, Elementary Form (Association Press)
Test of Religious Thinking, Advanced Form (Association Press)
Union Tests of Religious Ideas (Union Theological Seminary)
Union Tests of Ethical Judgment (Union Theological Seminary)
Whitley Biblical Knowledge Tests (M. T. Whitley)

Reputation Measures

Character Educational Inquiry Check Lists (Association Press)
Character Educational Inquiry Guess Who Tests (Association Press)
Chassell-Upton Citizenship Scales (Teachers College Bureau of Publications)
Check List of Traits (mimeographed) (Guidance Laboratory)
Haggerty-Olson-Wickman Behavior Rating Scale (World Book)

Environment Measures

Chapin: Measurement of Social Status (University of Minnesota)
McCall-Herring: Background Questionnaire (Laidlaw Bros.)
McCormick Scale for Measuring Social Adequacy (Catholic Univ. Press)
Minnesota Home Status Index (University of Minnesota Press)
Sims Score Card for Socio-Economic Status (Public School Publishing Co.)
Wallin Home Conditions, Personal and Family History Blank (Stoelting)
Whittier's Scale for Grading Neighborhood Conditions (California Bureau of Juvenile Research)

*Personality Tests**Elementary School Level:*

American Council on Education Rating Scale, Revised Form B,
(American Council on Education)

-
- Baker: Telling What I Do—Primary and Advanced Forms (Public School Publishing)
- Bregman Comprehensive Individual History Form (Psychological Corporation)
- Brown Personality Inventory for Children (Psychological Corporation)
- Burdick: Apperception Test (Association Press)
- Character Education Inquiry: Good Citizenship Test (Association Press)
- Character Education Inquiry: Good Manners Test (Association Press)
- Character Education Inquiry: Information Test (Association Press)
- Character Education Inquiry: Opinion Ballots A and B (Association Press)
- Character Education Inquiry: Portrait Matching Test (Association Press)
- Detroit Scale of Behavior Factors (Case Record) (Macmillan)
- Downy Individual Will Temperament Test (World Book)
- Downy Group Will Temperament Test (World Book)
- Hacker Character Rating Scale (McKnight & McKnight)
- Hayes: Scale for Evaluating School Behavior of Children Ten to Fifteen (Psychological Corporation)
- Hildreth Personality and Interest Inventory (Teachers College Bureau of Publications)
- Indiana Psychodiagnostic Blank (Indiana University)
- Kohs: Ethical Discrimination Test (Stoelting)
- Lehman: Play Quiz (Association Press)
- Maller Cooperation Tests (Association Press)
- Maller Character Sketches (Teachers College Bureau of Publications)
- Maller Self-Marking Tests (Teachers College Bureau of Publications)
- Maller Controlled Association Test (Teachers College Bureau of Publications)
- Maller Attention Test (Teachers College Bureau of Publications)
- Maller Character and Personality Scale (Teachers College Bureau of Publications)
- Maller Case Inventory (Teachers College Bureau of Publications)
- Maller Personality Sketches (Teachers College Bureau of Publications)
- New York Rating Scale for School Habits (World Book)
- Ohio State Personality Report Blank (Ohio State University)
- O'Reilly Character Analysis Chart (Public School Publishing Co.)
- Otis Suggestibility Test (Stoelting)
- Pintner, et al.: Pupil Portraits (Teachers College Bureau of Publications)
- Pressey Interest-Attitude Test (Psychological Corporation)
- Pressey X-O Tests (Stoelting)

Psychotic Questionnaire (Stoelting)
Rogers Test of Personality Adjustment (Form for girls and form for boys) (Association Press)
Rorschach Psychodiagnostic Test (Bircher)
Schwartz: Social Situation Pictures in the Psychiatric Interview (Stoelting)
Smith Self-Comparison Inventory (University of Minnesota)
Strang Test of Knowledge of Social Usage (Teachers College Bureau of Publications)
Tomlin: The Best Thing to Do (Stanford University)
Vineland Social Maturity Scale (Vineland Training School)
Wood: Right Conduct Test (Hillsdale School Supply Co.)
Woodworth Personal Data Sheet (Stoelting)

High School and College Level:

Allport A-S Reaction Study (Forms for men and forms for women) (Houghton Mifflin)
Allport: A Study of Personality—a systematic questionnaire (Stoelting)
Allport-Vernon Study of Values (Houghton Mifflin)
Almack: Sense of Humor Test (Gregory)
American Council on Education Rating Scale, Revised Form B. (American Council on Education)
Beckman: Revision of the A-S Reaction Study for Business Use (Houghton Mifflin)
Bell: Adjustment Inventory (Stanford University Press)
Bernreuter Personality Inventory (Stanford University Press)
Bregman Comprehensive Individual History Form (Psychological Corporation)
Brotmarkle: Comparison of Moral Concepts (Stoelting)
Character Education Inquiry Interest Analysis Test (Association Press)
Colegate Emotional Outlet Tests (Hamilton Republican)
Davis: Personal Problem Tests (Stoelting)
Detroit Scale of Behavior Factors (Case Record) (Macmillan)
Dougherty-O'Reilly Character Inventory Chart (Public School Publishing Co.)
Hayes: Scale for Evaluating School Behavior of Children Ten to Fifteen (Psychological Corporation)
Hildreth Personality and Interest Inventory (Teachers College Bureau of Publications)
Humm-Wadsworth Temperament Scale (Psychological Corporation)
Jones: Shall I Go to College (Public School Publishing Co.)
Kohs: Ethical Discrimination Test (Stoelting)
Lehman: Play Quiz (Association Press)
Loofbourrow-Keys: Personal Index (Educational Test Bureau)
MacNitt: A Psychological Interview (Psychological Corporation)

Maller: Character and Personality Rating Scale (Teachers College Bureau of Publications)
 Maller Character Sketches (Teachers College Bureau of Publications)
 Maller Self-Marking Test (Teachers College Bureau of Publications)
 Maller Case Inventory (Teachers College Bureau of Publications)
 Maller: Objective Test of Honesty (Teachers College Bureau of Publications)
 Maller Controlled Association Test (Teachers College Bureau of Publications)
 Maller Personality Sketches (Teachers College Bureau of Publications)
 Minnesota Personality Traits Rating Scales (Stoelting)
 Nebraska Personality Inventory (Sheridan Supply Co.)
 New York Rating Scale for School Habits (World Book)
 North Carolina Rating Scale for Fundamental Traits (Stoelting)
 Ohio State Personality Report Blank (Ohio State University)
 Otis Suggestibility Test (Stoelting)
 Pressey Sports Information Test (Stoelting)
 Pressey Interest-Attitude Test (Psychological Corporation)
 Pupil Portraits (Teachers College Bureau of Publications)
 Psychotic Questionnaire (Stoelting)
 Root: Introversion-Extroversion Test (Psychological Corporation)
 Rorschach Psychodiagnostic Test (Bircher)
 Smith: Self-Comparison Inventory (University of Minnesota)
 Strang: Test of Knowledge of Social Usage (Teachers College Bureau of Publications)
 Stephenson-Millet: Test on Social Usage (McKnight and McKnight)
 Symonds: Adjustment Questionnaire (Psychological Corporation)
 Symonds: Student Questionnaire (Teachers College Bureau of Publications)
 Symonds: What Kind of a Year Are You Having? (Teachers College Bureau of Publications)
 Thurstone Personality Schedule (University of Chicago Press)
 Vineland Social Maturity Scale (Vineland Training School)
 Wechsler Self-Administering Maze (Psychological Corporation)
 Washburne Test on Social Adjustment (Washburne)
 Willoughby Emotional Maturity Scale (Stanford University Press)
 Tests of the Socially Competent Person (Teachers College Bureau of Publications)

Curriculum

McCall-Herring-Loftus: School Practices Questionnaire (Laidlaw)

Sewing Tests

Murdoch Sewing Scale (Teachers College Bureau of Publications)
 Murdoch Analytic Sewing Scale for Measuring Separate Stitches (Teachers College Bureau of Publications)

*Tests and Rating Scales in Education**For College Students, Teachers, Supervisors, and Principals:*

- Almy-Sorenson Rating Scale for Teachers (Public School Publishing Co.)
- Bathurst-Knight-Ruch-Telford: Aptitude Test for Elementary and High School Teachers (Bureau of Public Personnel Administration)
- Bathurst-Knight-Ruch-Telford: Placement Test for Elementary Teachers (Bureau of Public Personnel Administration)
- Brown: A Self-Rating Scale for Supervisors, Supervisory-Principals, and Helping Teachers (Bruce Publishing Co.)
- Brueckner: Judgment Test of Teaching Skill (Educational Test Bureau)
- Carrigan Score Card for Rating Teaching and the Teacher (World Book Co.)
- Coöperative Professional Education Test (Educational Records Bureau)
- Coxe-Orleans Prognosis Test of Teaching Ability (World Book Co.)
- Edmondson-Schorling: Practical Problems in Education (Public School Publishing Co.)
- Frasier-Armentrout: An Introduction to Education (Scott, Foresman)
- George Washington University Teaching Aptitude Test (Center for Psychological Service)
- Geyer: Objective Examination on Intelligence Testing (Plymouth Press)
- Howe-Kyte: Diagnostic Record of Teaching (Houghton Mifflin)
- Johnson Checking List and Standards for Supervision of High School Instruction (Teachers College Bureau of Publications)
- Jordan: Objective Tests on Educational Psychology (Holt)
- Kefauver-Hand Educational Guidance Tests (World Book)
- Lewerenz-Steinmetz: Orientation Test Concerning Fundamental Aims of Education (Southern California School Book Depository)
- Michigan Education Association Teacher Self-Rating Scale for Self-Improvement (Michigan Education Association)
- Minnesota Rating Scale for Teachers of Home Economics (University of Minnesota Press)
- Odell-Herriott: Standard Achievement Test in Principles of Teaching (Public School Publishing Co.)
- Peik: Recitation Analysis and Survey Check List (Educational Test Bureau)
- Potthoff-Corey: Tests in Educational Psychology (Public School Publishing Co.)
- Schutte: Scale for Rating Teachers (World Book)
- Stanford Educational Aptitude Test (Stanford University Press)
- Van Hoesen: Comprehensive Examination in Education (Ann Arbor Press)

-
- Van Wagenen Reading Scales in Educational Psychology (Educational Test Bureau)
Votaw-Grinnel: Standardized Examination on Principles of Secondary Education (Harlow Publishing Co.)
Waples and Reavis: Classroom Procedure Test (University of Chicago Press)
Weber: Standard Achievement Test on Aims, Purposes, Objectives, Attributes, and Functions in Secondary Education (Public School Publishing Co.)
Woodworth: Revised Standardized Examination in Psychology (Henry Holt)

Laboratory Tests:

- Briggs Analogies Test (Teachers College Bureau of Publications)
Burt Graded Reasoning Test (mimeographed) (Guidance Laboratory)
Coördination Test (Association Press)
Directions Tests I, II, III, and IV (Stoelting)
Goddard Memory Picture Test (Stoelting)
Plymouth Educational Miscellaneous Laboratory Tests (Plymouth Press)
Ruggles Distraction Test (Association Press)
Speed Tests (Association Press)
Thurstone Substitution Test (Stoelting)
Whipple Analogies Test (Stoelting)

Mechanical Ability Tests:

- Badger: Mechanical Drawing Test (Public School Publishing Co.)
Castle: Mechanical Drawing Test (Manual Arts Press)
Detroit Mechanical Aptitude Tests (Forms for girls and forms for boys) (Public School Publishing Co.)
Institute of Educational Research Paper Cutting Test (Stoelting)
Kelley Construction Ability Test (Stoelting)
MacQuarrie Test for Mechanical Ability (Southern California School Book Depository)
Minnesota Mechanical Ability Tests (Stoelting)
Minnesota Paper Form Board Test (University of Minnesota)
Minnesota Spatial Relations Test (Stoelting)
Minnesota Rate of Manipulation Test (University of Minnesota)
Newkirk-Stoddard Home Mechanics Test (State University of Iowa)
Stenquist Mechanical Ability Tests (Stoelting)
Stenquist Mechanical Aptitude Test (Stoelting)
Wells-Lubach: Mechanical Drawing Test (Manual Arts Press)
Wiggley Block Test (Stoelting Company)
Wright Achievement Test in Mechanical Drawing (Public School Publishing Co.)

Vocational Tests and Interest Scales

- A-B-C Occupational Inventory (Publication Press)
Aids to Vocational Interview, Record Form B (Psychological Corporation)
Brainard Specific Inventories (Forms for Men, Women, Boys, and Girls) (Psychological Corporation)
Cleeton Vocational Interest Inventory (Forms for Men and Women) (Psychological Corporation)
Cleeton-Mason Vocational Aptitude Examination (Psychological Corporation)
Comprehensive Objective Examination in Commercial Law (Harlow)
Crabbe-Slinker: Achievement Test for General Business Training (South-Western Publishing Co.)
Dillavou-Greiner: Business and Law Objective Test (McGraw-Hill)
Ferson-Goddard Law Aptitude Examination (West Publishing Co.)
Freyd: Occupational Interest Blank for Men (Stoelting)
Garretson-Symonds: Interest Questionnaire for High School Students (Teachers College Bureau of Publications)
George Washington University Scholastic Aptitude Test for Medical Students (Center for Psychological Service)
George Washington University Test for Automobile Drivers (Center for Psychological Service)
George Washington University Test for Ability to Sell (Center for Psychological Service)
George Washington University Aptitude Test for Nursing (Center for Psychological Service)
Hepner Vocational Interest Quotient Booklets (Psychological Corporation)
Hoppock Questionnaire for Studies of Job Satisfaction (Robert Hoppock)
Kefauver-Hand Vocational Guidance Test (World Book)
Leonard Rating Scale for Predicting Success (Houghton Mifflin)
Leahy-Fenelson Rating Scale for Social Case Workers (University of Minnesota Press)
Lufburrow Vocational Interest Locator (Publication Press)
McHale Vocational Interest Test for College Women (American Association of University Women)
Manson Occupational Interest Blank for Women (Psychological Corporation)
Miner Analysis of Work Interests (Stoelting)
Minnesota Interest Analysis Blank (University of Minnesota)
Minnesota Rating Scale for Social Case Workers (University of Minnesota Press)
Morris Trait Index (Public School Publishing Co.)
Ohio State Educational and Vocational Information Blank (Ohio State University)
Ohio State Educational Intentions Blank (Ohio State University)

-
- Ohio State Vocational Information Blank (Ohio State University)
 Otis General Intelligence Examination for Business Institutions
 (World Book)
 Parke Commercial Law Test (Kansas State Teachers College)
 Personnel Counseling Service Blanks (Cooperative Counseling Service)
 Personnel Research Federation Personal History Record (Bureau of
 Personnel Research, Personnel Research Federation)
 Prosser-Anderson: Practice Book on Getting a Job (McKnight &
 McKnight)
 Public Personnel Administration Test for Automobile Mechanics
 (Bureau of Public Personnel Administration)
 Record of Proficiency in Nursing Practice (University of Minnesota
 Press)
 S.O.G.I. Interest Scale (Guidance Laboratory)
 Sondquist Interest Finder (Association Press)
 Steno-Gauge Test (Psychological Corporation)
 Strong Vocational Interest Blank for Women (Stanford University
 Press)
 Strong Vocational Interest Blank for Men (Stanford University
 Press)
 Stuart Objective Tests in Typewriting (Gregg Publishing Co.)
 Teeter Objective Tests in Guidance (McGraw-Hill)
 Thompson Business Practice Test (World Book)
 Thurstone Vocational Guidance Test in Algebra (World Book)
 Thurstone Vocational Guidance Test in Arithmetic (World Book)
 Thurstone Vocational Guidance Test in Geometry (World Book)
 Thurstone Vocational Guidance Test in Physics (World Book)
 Thurstone Vocational Guidance Test in Technical Information
 (World Book)
 Thurstone Vocational Interest Schedule (Psychological Corpora-
 tion)
 Westin Commercial Law Achievement Test (Southern California
 School Book Depository)
- Clerical Ability:*
- Detroit Clerical Aptitudes Examinations (Public School Publishing
 Co.)
 Graphic Rating Scale for Clerical Workers (mimeographed) (Guid-
 ance Laboratory)
 Linke and Koehne: Topical Filing Test (Stoelting)
 Minnesota Vocational Test for Clerical Workers (Psychological
 Corporation)
 O'Rourke Clerical Aptitude Test (Educational and Personnel Pub-
 lishing Co.)
 Scott Filing Test (Stoelting)
 Stalnaker: Examination in Clerical Proficiency (Psychological
 Corporation)
 Thurstone Employment Test in Clerical Work (World Book)

Bookkeeping:

- Altholz-Braverman: Modern Bookkeeping Practice Tests (Lyons and Carnahan)
Baker-Prickett-Carlson: Bookkeeping Tests (South-Western Publishing Co.)
Bowman Bookkeeping Achievement Test (American Book Co.)
Bredenbaugh Bookkeeping Tests (Public School Publishing Co.)
Carlson: Bookkeeping Tests (South-Western Publishing Co.)
Detroit Bookkeeping Examination (Board of Education, Detroit)
Ellwell-Fowlkes Bookkeeping Test (World Book)
Jackson-Sanders-Sproul: Bookkeeping Tests (Ginn & Co.)
Studebaker, et al.: Bookkeeping Test (Purdue University)

Stenography:

- Bisbee Shorthand Test (Public School Publishing Co.)
Blackstone Stenographic Proficiency Test (World Book)
Comprehensive Objective Examination in Gregg Shorthand (Harlow)
Detroit Shorthand Examination (Board of Education, Detroit)
Hoke: Prognostic Test of Stenographic Ability (Gregg Publishing Co.)
Hoke: Tests in Gregg Shorthand (Gregg Publishing Co.)
Rollinson Diagnostic Shorthand Test (Psychological Corporation)

Typewriting:

- Clem Typewriting Test (Public School Publishing Co.)
Kauzer Typewriting Test (Kansas State Teachers College)
North Objective Tests for Teachers of Typewriting (Gregg Publishing Co.)
Stalnaker: Test of Typewriting Ability (Psychological Corporation)
Stuart Objective Tests in Typewriting (Gregg Publishing Co.)
Thurstone Employment Tests in Typewriting (World Book)

Performance Test Material

- Atkins Object-Fitting Test (Stoelting)
Brace Scale of Motor Ability Tests (Barnes & Co.)
Color Patterns Test—Yerkes Point Scale (Stoelting)
Cornell-Coxe Performance Ability Scale (World Book Co.)
Dearborn Form Board #3 (Stoelting)
Dearborn Form Board #4 (Stoelting)
Dearborn Reconstruction Board (Stoelting)
Dunham Arrow Board (Stoelting)
Ferguson Form Boards (Stoelting)
Goddard Adaptation Board (Stoelting)
Gwyn Triangle (Stoelting)
Healy Construction Board B (Stoelting)
Healy Fernald Construction Puzzle A (Stoelting)

I.E.R. Assembly Test for Girls (Stoelting)
 Kempes Diagonal Form Board (Stoelting)
 Knox Cube Imitation Test (Stoelting)
 Knox Moron Test (Stoelting)
 Kohs Block Design (Stoelting)
 Learning Test Cubes (Stoelting)
 Maxfield Color Cubes (Stoelting)
 Merrill-Palmer Test (R. Stutsman)
 Minnesota Assembly Test, A—B—C (Stoelting)
 Minnesota Assembly Boxes I and II (Stoelting)
 Minnesota Card Sorting Test (Stoelting)
 Minnesota Packing Blocks Test (Stoelting)
 Minnesota Pre-School Test (Educational Test Bureau)
 Minnesota Rate of Manipulation Test (University of Minnesota)
 Minnesota Spacial Relations Test (University of Minnesota)
 Mullan Memory of Objects Test (Stoelting)
 Otis Test of Suggestibility (cards for test) (Stoelting)
 Passalong Test (J. and J. Cook)
 Pintner-Patterson Performance Test (Stoelting)
 Porteus Form and Assembly Test (cards for test) (Vineland Training School)
 Rossolimo Test (Stoelting)
 Seguin Form Board (Goddard Modification) (Stoelting)
 Seguin-Witmer Sylvester Form Board (Stoelting)
 Slot-Young Maze, Test A (Stoelting)
 Stenquist Mechanical Tests I, II, III (Stoelting)
 Town Picture Memory Test (Stoelting)
 Wallin Peg Boards (Stoelting)
 Witmer Cylinder Test (Stoelting)
 Wooley-Fischer Immediate Memory Test (Stoelting)

List of Publishers

Alabama Polytechnic Institute Auburn, Alabama	Ann Arbor Press Ann Arbor, Michigan
American Assoc. of University Women 106 East 52nd Street New York City	Association Press 347 Madison Avenue New York City
American Book Company 88 Lexington Avenue, New York City	Avent, Joseph E. Box 1455, Knoxville, Tennessee
American Council on Education 744 Jackson Street Washington, D. C.	Barnes, A. S. Co. 67 West 44th Street New York City

Bircher, Ernest Verlag Bern Und Leipzig, Germany	Clatworthy, L. M. University of Denver Denver, Colorado
Bruce Publishing Co. 40 East 34th Street New York City	Clio Press Iowa City, Iowa
Buffalo, University of Buffalo, New York	College Book Company Columbus, Ohio
Bureau Educational Research, Cleveland Public Schools, Cleveland, Ohio	Columbia Test Service 855 North Nelson Road Columbus, Ohio
California Bureau of Juvenile Research Whittier State School Whittier, Calif.	Cook, J. and J. Paisley, Glasgow, Scotland
Catholic University of America Washington, D. C.	Coöperative Counseling Service 715 South Hope Street Los Angeles, Calif.
Center for Psychological Service 2026 G Street, N. W. Washington, D. C.	Coöperative Test Bureau 347 West 59th Street New York City
Century Publishing Company 353 Fourth Avenue New York City	Cottrell, Dr. Donald Teachers College, Columbia Uni- versity New York City
Character Research Institute Washington University St. Louis, Missouri	Courtis, Dr. S. A. 1807 East Grand Boulevard Detroit, Michigan
Chassell, Dr. J. O. University of Rochester Rochester, N. Y.	Detroit, Board of Education Detroit, Michigan
Chicago Press University of Chicago Chicago, Ill.	Doubleday, Page and Company Garden City, Long Island, N. Y.
Cincinnati, University of Cincinnati, Ohio	Durrell, Dr. Donald Boston University Boston, Mass.
	Education and Personnel Pub- lishing Co. Washington, D. C.

Educational Records Bureau 437 West 59th Street New York City	Harlow Publishing Co. 217 North Harvey Street Oklahoma City, Oklahoma
Educational Test Bureau 3416 Walnut Street Philadelphia, Penna.	Harrap & Company, Ltd., George G. 39-41 Parker Street Kingsway, London, W. C. 2, England
Farley, Eugene S. Director of Research Newark, N. J.	Harter School Supply Co. Cleveland, Ohio
Fischer Company 56 Cooper Square New York City	Harvard University Press Cambridge, Mass.
Follett Publishing Company 1257 South Wabash Avenue Chicago, Illinois	Heath and Company, D. C. 180 Varick Street New York City
Friendship Press 150 Fifth Ave. New York City	Hildreth, Dr. Gertrude Lincoln School 425 West 123rd Street New York City
Ginn and Company 70 Fifth Avenue New York City	Henry Holt and Company 1 Park Avenue New York City
Gregg Publishing Company 270 Madison Avenue New York City	Hillsdale School Supply Co. 39 North Street Hillsdale, Michigan
Gregory Co., C. A. 347 Calhoun Street Cincinnati, Ohio	Hoppock, Robert Teachers College, Columbia Uni- versity New York City
Hamilton Republican Hamilton, New York	Houghton Mifflin Co. 386 Fourth Avenue New York City
Handschin, C. H. Miami University Oxford, Ohio	Illinois, University of Urbana, Illinois
Harcourt, Brace & Co. 383 Madison Avenue New York City	Independent Press Mexico, New York

Indiana, University of Department of Psychology Bloomington, Indiana	McGraw-Hill Book Company 330 West 42nd Street New York City
Indiana University Bookstore Indiana University Bloomington, Indiana	McKnight and McKnight Bloomington, Illinois
Iowa, State University of Bureau of Educational Research and Service Iowa City, Iowa	Macmillan Company 60 Fifth Avenue New York City
Johns Hopkins Press Baltimore, Maryland	Madison, State Department of Public Instruction Madison, Wisconsin
Jones, L. R. 227-9 E. Fourth Street Los Angeles, Cal.	Maller, Dr. Julius B. Teachers College, Columbia Uni- versity New York City
Kansas State Teachers College Bureau of Educational Measure- ments Emporia, Kansas	Mentzer, Bush and Company 55 Fifth Avenue New York City
Keystone View Company Meadville, Penna.	Michigan Education Association Lansing, Michigan
Lafayette Printing Company Lafayette, Indiana	Michigan, University of Department of Education Ann Arbor, Michigan
Laidlaw Brothers, Inc. 320 E. 21 Street Chicago, Ill.	Minnesota Press, Univ. of Minneapolis, Minnesota
Landis, R. H. Eastern Illinois State T. C. Charleston, Illinois	Missouri, University of Columbia, Missouri
Lippincott and Company 227 South 6th Street Philadelphia, Penna.	Mount Holyoke College South Hadley, Mass.
Little, Brown and Company 60 East 42nd Street New York City	New Mexico, University of Albuquerque, New Mexico
	Newson and Company 73 Fifth Avenue New York City

New York City, Board of Education 500 Park Avenue New York City	Public Personnel Administration, Bureau of Box 226 Trenton, New Jersey
North Carolina, University of Chapel Hill, North Carolina	Public School Publishing Co. Bloomington, Illinois
Northwestern High School Detroit, Michigan	Publication Press 1511 Guilford Avenue Baltimore, Maryland
Ohio State Dept. of Education Columbus, Ohio	Purdue University Research Foundation Lafayette, Indiana
Ohio State University Columbus, Ohio	Raup, Dr. R. B. Teachers College, Columbia University New York City
Palmer Co., A. N. 55 Fifth Avenue New York City	Row, Peterson & Co. 131 East 23rd Street New York City
Philadelphia, Bd. of Education Philadelphia, Penna.	Russell Sage Foundation 130 East 22nd Street New York City
Pintner, Dr. R. Teachers College, Columbia University New York City	Saskatchewan, University of Saskatoon, Saskatchewan, Canada
Plymouth Press 6749 Wentworth Avenue Chicago, Illinois	Scott, Foresman and Company 114 East 23rd Street New York City
Practical Arts Publishing Co. 44 Vista Avenue Elizabeth, N. J.	Scribner's Sons, Charles 597 Fifth Avenue New York City
Price, E. D. Enid, Oklahoma	Sheridan Supply Co. P. O. Box 1009 Lincoln, Nebraska
Psychological Corporation 522 Fifth Avenue New York City	Shields, F. J. 1594 Whitefield Road Pasadena, California

BOOK THREE

USE OF STANDARD TESTS FOR
GROUPING PUPILS

CHAPTER VIII

HOW TO ADMINISTER TESTS AND OBTAIN AND INTERPRET GRADE OR AGE SCORES¹

1. HOW TO ADMINISTER TESTS

Best Dates for Testing.—Perhaps the best time to administer standardized tests is at the beginning of a term. Another good time to administer standardized tests is at the end of the school term—June in the annual school, or January and June in a semi-annual school. An advantage of applying standardized tests at the end of the term is that, in addition to supplying information which will be useful in the promotion of the pupils, the tests may be substituted for the traditional teachers' examinations given at those times, not only reducing thus the amount of work required of the teacher, but also supplanting the traditional tests by easily scorable, significant measures of pupil ability. After the tests have been selected and purchased, it is necessary to decide who shall administer them to the pupils. In most schools this is done by the classroom teacher. To insure uniformity, however, it is often desirable to have all the tests in a school given by one person, such as the assistant principal or a teacher who has taken special training in measurement.

If the classroom teachers are to give the tests, they should be carefully trained in test procedure. The following program is recommended:

1. At least one member of the faculty should take a course in measurement.
2. This person, or a trained examiner, such as the director of research, should administer one or more tests in the presence of each teacher.
3. The teacher's attention should be called to the essential points to be observed in administering tests, as outlined below.
4. The teacher should administer at least one test under the personal supervision of the trained examiner. The examiner will assist the teacher to perfect her technique in every respect.

¹ The author makes grateful acknowledgment of Dr. Harold H. Bixler's help in preparing portions of Book Three.

5. The teachers may exchange classes for the administration and scoring of tests, although each teacher should rescore for his own pupils.

Points to Be Observed.—In administering a standard test, the examiner should keep in mind the following points:

1. *Make sure that he knows how to administer the test.*—He should be thoroughly familiar with the manual of directions and with the test itself.

2. *Insure good working conditions.*—The test will usually be given in the classroom, but occasionally it will be necessary to assemble the pupils in the auditorium. It should be remembered, however, that an unusual environment may be distracting to some children. Pupils should be comfortably seated with a desk or other convenient place on which to write. Both lighting and temperature should be correct. Pupils should be seated far enough apart to lessen opportunities for intentional or unintentional copying.

3. *Anticipate and avoid possible distractions.*—A notice should be placed on the door to warn teachers, messengers, janitors, and visitors not to enter while the test is in progress. Distractions on the outside, such as noise on the playground or in the street, should be eliminated if possible. Window shades should be adjusted so that they do not flap.

4. *Have all necessary material at hand.*—This will include the manual of directions, the correct number of test papers, a watch with a second hand, and a few extra pencils with erasers.

5. *Secure attention before beginning.*—In many tests, directions are stated only once. An inattentive pupil is at a serious disadvantage.

6. *Put the pupils in the proper attitude of mind.*—The examiner should strive to prevent nervousness or tension, and to put the children at their ease. This should be done in a pleasant manner and by general remarks that have nothing to do with the test. Any statement about the test or any motivation for the test not provided in the standard directions or instructions should be carefully avoided.

7. *See that all pupils have pencils and, if possible, erasers.*—The examiner should inform the pupils that a supply of pencils is available on his desk, in case any pupil has only one pencil and breaks the point.

8. *Instruct pupils what to do when papers are distributed.*—For example, the direction in the Multi-Mental Test is, "Please leave them face down until I tell you to turn them."

9. *Distribute papers.*—The examiner or an assistant should give a sufficient supply to pupils in the front row, and have them distribute the tests to their sections. They should be informed whether they are to place the papers face down or up.

10. *Obtain the necessary identifying information.*—The examiner must give pupils time to write their names and other desired information. With very young children, the examiner may need to give assistance, or to have names placed on the papers in advance.

11. *Read the directions to pupils verbatim.*—The examiner will ordinarily not memorize the directions nor will he paraphrase them. He must read slowly and distinctly. No remarks or explanations should be added unless the manual so directs.

12. *Keep time accurately.*—Both the starting time and the time when the pupils are to be stopped should be recorded to the second; thus, 9 : 40 : 13. The form of expression, "Twenty minutes before ten," should be avoided.

13. *Permit no questions aloud after the pupils have begun work.*—Finger on lips at the first question will cause pupils to raise a hand instead of speaking. The examiner should go quietly to any pupil who has raised his hand, to see what the trouble is.

14. *Give no assistance on the test items proper but aid any pupil who has not understood directions as to the mechanics of taking the test.*—The only exceptions to this point are certain intelligence tests where the test consists in the pupil's ability to follow verbal directions.

15. *Just after the test begins and occasionally thereafter move quietly about the room to see that identification blanks have been filled correctly and that pupils are proceeding in accordance with the directions for the test.*—A good position when not moving about the room is at one of the front corners. The examiner must be on the alert for unusual occurrences, such as an imperfect test blank, cheating, and the like.

16. *Collect all papers at one time.*—It is usually not desirable to permit pupils to hand in their papers as they finish.

17. *Obtain all papers.*—Be sure that no used or unused tests are left in the hands of pupils or in any place where they might fall into the hands of pupils.

18. *Strictly follow the published standard directions for each test, even though they conflict with the general directions given above.*—Standard directions always have priority.

2. HOW TO OBTAIN CRUDE SCORES

General Suggestions.—The manuals of directions which accompany the tests contain instructions for scoring. A few general suggestions as to procedure are given here.

1. Preferably teachers should score all tests. Clerks or older pupils may be used. In certain instances, if properly supervised, pupils may exchange papers and score each other.

2. Use the key or stencil provided with the test. If there is none, a key can be made by taking a test paper and writing the correct answers in red. Place this key paper beside the paper to be scored, close to the pupil's answers, and the answers can be quickly compared.

3. It is better to score consecutively the same subtest or section of a test in all papers than to score each complete paper separately. If there are a number of scorers, each should specialize on one section of the test, passing the papers from one to another as each scorer finishes his assignment. When there are several scorers, each one should write his initials on the page or section scored, so that if errors are discovered they can be traced to the person responsible.

4. Adopt a uniform plan of scoring; for example, mark rights with a dash (—) and wrongs with a zero (0). These marks are recommended because they can be made quickly with one stroke of the pencil. Checking in this way facilitates counting.

5. In case of doubt, try to ascertain the intent of the pupil. If there is no evidence of correct intent, score as wrong.

6. Record the score in the space provided; if none is provided, use the upper right-hand corner of the test paper.

7. All papers should be rescored or checked by another scorer, or by the same scorer at a later time. If it is impossible to rescore all papers, be sure to check a random sample at least. After some experience in checking for errors, certain sections of a test or certain subtests will be discovered in which scorers show a constant tendency to make errors. Such sections should be checked in all papers. The checker should write his initials on the papers checked.

The Crude Score.—The crude or raw score is the total score obtained on the test. If this total score is obtained by adding scores on subtests or parts of a test, these arithmetical computations should be checked.

The total score is known as a crude score to distinguish it from derived scores, such as the grade score, which is described in the next section.

3. HOW TO OBTAIN GRADE SCORES

What the Grade Score Is.—The next step is to convert crude scores into grade scores (commonly abbreviated to G scores). The grade score expresses the ability or achievement of a pupil in terms of the ability or achievement of an average of a large number of pupils of a given grade, and thus indicates approximately his appropriate grade classification in a typical school.

Tables for finding G and age scores are usually supplied with tests. In this chapter, for purposes of illustration, the tables for the *Thorndike-McCall Reading Scale* and the *Woody-McCall Mixed Fundamentals in Arithmetic* are used. More recent tests might have been used, but data for these tests were available, and had been worked up and carefully checked for accuracy. The procedure is the same regardless of the tests used.

Part of the G and age score table for the *Reading Scale* is reproduced in Table 2. It is read as follows: First find the column corresponding to the Form of the test that is used. For example, if Form 5 is used, all work will be done with the column headed Form 5. Suppose John has a crude score of 3. Find 3 in the column headed Form 5; then, in the column headed G score, read the figures opposite 3. We find 2.6. A crude score of 3 on the *Reading Scale*, therefore, may be transmuted into a grade score of 2.6 (usually read "two point six"). A grade score of 2.6 means that the achievement of the pupil is equivalent to that of the average second grade pupil after .6 of a school year of instruction.

A crude score of 3 converts to an age score 7.5, i.e., the pupil reads as well as the typical pupil who is $7\frac{1}{2}$ years of age.

The G and age score table for the *Woody-McCall Mixed Fundamentals in Arithmetic Scale*, part of which is reproduced in Table 3, is used in the same way. If, for example, Form II is used and John has a crude score of 10, we find 10 in the column

headed Form II and read the corresponding G score. We find it to be 3.4.

The crude score of 10 converts to an age score of 8.2.

The G and age table for the *Multi-Mental Scale* is found in the manual of directions for that test.

Similarly, G and age scores in spelling appear in the manual of directions for the *Morrison-McCall Spelling Scale*.

The method of securing G or age scores in the case of other standard tests is similar. The Public School Publishing Company, however, provides "B" score tables for many of the tests which they publish. Their B scores are identical with G scores.

In the manuals of directions for *Gates Primary Reading Tests* and *Gates Silent Reading Tests*, the term Reading Grade is used. Reading grade scores are the same as G scores in reading.

For some standard tests G and age tables are not provided. In Book Seven a technique for constructing tables is described.

Labeling G Scores.—The various grade scores are labeled so that we may distinguish them. They are written as follows:

Grade Score in Intelligence.....	Gi
Grade Score in Reading.	Gr
Grade Score in Arithmetic.....	Ga
Grade Score in Spelling	Gs
Grade Score in Handwriting.....	Gha
Grade Score in History.....	Ghi
Grade Score in Education	Ge
Grade Score for Placement.....	Gp

It is of course possible to have G scores in other subjects. They are designated by combining the initial letter or letters of the subject with G.

If two or more tests are used in any subject, they are differentiated by adding numerals to the label. For example, one reading test may be labeled Gr1; another reading test, Gr2.

4. HOW TO INTERPRET GRADE AND AGE SCORES

Use of G or Age Scores.—Grade or age scores are much more useful and practical than crude scores because they are more easily interpreted and because by means of them scores on all tests can be put in comparable units. In Table 4, for example, Pupil 1 has a Gi of 3.4. This signifies intelligence equal to the

TABLE 2
G TABLE FOR THORNDIKE-McCALL READING SCALES

AGE SCORE	CRUDE SCORE							G SCORE
	Form 1	Form 2	Form 3	Form 4	Form 5	Form 6	Form 7	
5.9	—	—	—	—	—	—	0	1.4
6.0	0	—	—	—	—	—	1	1.5
6.2	—	0	0	0	—	—	2	1.7
6.3	1	—	—	—	0	0	3	1.8
6.5	—	1	1	—	—	—	4	2.0
6.8	2	2	2	—	—	—	5	2.1
7.0	3	3-4	3	1	1	1	6	2.3
7.3	4	5	4	2	2	2	7	2.4
7.5	5-6	6	5	3-4	3	3	8-9	2.6
7.8	7	—	6	5	4	4	10	2.8
8.0	8-9	7	7	6-7	5-6	5	11	2.9
8.3	10	8	8	8	7	6	12	3.1
8.4	11	9	9-10	9-10	8-9	7-8	13	3.3
8.7	12	10	11	11	10-11	9-10	14	3.5
8.9	13	11	12	12	12-13	11-12	15	3.7
9.2	14	12	13	13	14	13	—	3.8
9.4	15	13	14-15	14-15	15	14	16	4.0
9.7	16	14	16	16	—	15	—	4.2
9.8	—	15	17	—	16	16	17	4.3
10.1	17	16	18	17	17	17	—	4.5
10.3	18	—	—	18	18	—	18	4.7
10.5	—	17	19	19	—	18	19	4.9
10.8	19	18	20	—	19	19	—	5.1
11.1	—	—	—	20	20	20	20	5.3
11.3	20	19	21	—	21	—	—	5.4
11.5	—	—	22	21	—	21	21	5.6
11.8	21	20	—	—	22	—	—	5.8
12.1	—	—	23	22	23	22	22	6.0
12.3	22	21	—	—	—	—	—	6.1
12.5	—	—	24	—	24	23	23	6.3
12.7	23	22	25	23	—	—	—	6.5
13.0	—	—	—	—	25	24	24	6.7
13.2	24	23	—	24	—	—	—	6.8
13.4	—	—	26	—	26	25	25	7.0
13.7	25	—	27	25	—	—	26	7.2
13.9	—	24	—	26	27	26	—	7.4
14.1	26	—	—	—	—	—	27	7.7
14.4	—	25	28	27	—	—	—	8.0
14.6	27	—	—	—	28	27	28	8.3
14.9	—	26	29	—	—	—	—	8.8
15.1	28	—	—	28	29	28	29	9.2

TABLE 3
G TABLE FOR MIXED FUNDAMENTALS IN ARITHMETIC SCALES ¹

AGE SCORE	CRUDE SCORE				G SCORE
	Form I	Form II	Form III	Form IV	
6.8	—	0	—	—	2.0
6.8	0	1	—	—	2.0
6.9	1	2	—	0	2.1
7.1	2	3	—	1	2.3
7.2	3	4	—	—	2.4
7.3	4	—	0	2	2.6
7.4	5	5	—	—	2.7
7.6	6	6	—	—	3.0
7.7	7	7	1	3	3.1
7.8	8	8	2	—	3.2
8.0	—	9	3	4	3.3
8.0	9	—	—	5	3.3
8.2	—	10	4	6, 7	3.4
8.3	—	11	5, 6	8	3.5
8.5	10	12	7	9	3.6
8.6	11	—	8	10	3.7
8.7	12	13	9	—	3.9
8.8	13	14	10	11	4.0
9.0	14	15	11	12	4.2
9.2	15	16	12	13	4.5
9.3	16	17	13	14	4.7
9.4	—	—	14	—	4.8
9.6	17	18	—	15	4.9
9.7	18	19	15	16	5.2
10.0	19	—	16	17	5.3
10.2	—	20	17	18	5.4
10.3	20	21	—	—	5.6
10.5	21	—	18	19	5.8
10.7	22	22	19	20	6.0
10.9	—	23	—	—	6.1
11.2	23	24	20	21	6.2
11.3	—	—	21	22	6.4
11.5	24	25	22	—	6.5
11.8	—	26	—	23	6.7
12.0	25	27	23	24	6.8
12.5	26	—	24	25	7.2
12.8	27	—	25	—	7.5
13.0	—	28	—	26	7.7
13.3	—	—	26	—	7.9
13.6	28	29	—	27	8.0

¹ Calibrated by the Bureau of Research, Board of Education, New York City and Bureau of Educational Research, University of Michigan.

average third-grade pupil after four months of instruction. The Gr for Pupil 1 is 2.1, which means that his reading ability is equivalent to that of the average second-grade pupil after one month of instruction. In general, then, the figure of a G score to the left of the decimal point indicates the grade level. The figure to the right of the decimal point indicates the number of months at this grade level.

A Gr of 2.1 may also be read two and one-tenth, being interpreted to mean that the pupil's achievement is equivalent to that normally accomplished in two and one-tenth grades. Where the school year is shorter than ten months, one should proceed on the ten-months basis, considering each month as one-tenth of the school year. (See Section 6, "The School with a Short Term," in Chapter XIII.)

The interpretation of age scores is so obvious as to require no discussion.

CHAPTER IX

HOW TO COMBINE GRADE OR AGE SCORES

In the preceding chapters we have explained how to administer tests, how to obtain crude scores, and how to convert crude scores into G or age scores. This chapter will explain how to combine two or more G scores in the same subject. The procedure is similar for age scores.

It may happen that two reading tests or two or more arithmetic tests have been used to secure greater reliability or for diagnostic purposes. These G scores are usually differentiated by means of a number; in the case of reading, for example, they would be written Gr1, Gr2, etc. These scores may be recorded separately on the Class Record Sheet if diagnosis and analysis of test scores are desired (see Table 17 B). For the sake of brevity, however, it may be desirable to combine these separate scores into a single figure (see Table 17 C). The Class Record Sheet will then contain only one Gr or grade score in reading, and one Ga or grade score in arithmetic, etc., for each pupil. Comparisons such as between Gr and Gi, Ga and Gi, or Gr and Ga are thus facilitated. The technique for combining two or more such G scores in a single subject is given below.

Step 1. Determine What Weights to Use.—The first step is to determine how the tests shall be weighted. Weights refer to the relative values to be assigned to tests when combining G scores. For example, if two reading tests are given equal weight, their G scores would be averaged. On the other hand, if it is desired to give one test twice as much weight as another, the score on the test so valued would be multiplied by 2.

The examiner will determine the weights to be used on the basis of the significance of the tests for his purpose. The significance of a test depends upon the following factors:

(1) *The trait measured.*—For example, one reading test may measure recognition of words; another, comprehension of sentences and paragraphs. If the pupils to be classified are in the first grade, these traits might be judged to be of equal value; if

they are in the second grade, more weight might be given to the second trait, comprehension.

(2) *The reliability or accuracy of the test.*—This is usually reported by the author of the test. A test is reliable when it measures accurately, that is, when application of equivalent forms yield scores that are practically identical. Other things being equal, the greater the number of test items, the more reliable the tests. A rough index of reliability is the working time, that is, the number of minutes required to take the test. It must be understood, however, that working time is only a rough index of reliability.¹ One must be especially careful in the case of tests with a liberal time limit,—when pupils actually work only part of the time allowed. No arbitrary rule can be set up for determining weights on the basis of reliability. The examiner must use his discretion.

A battery of tests recommended for primary grades is the group of *Gates Primary Reading Tests*. The time limits for these tests are:

Type 1. Word Recognition. 15 minutes.

Type 2. Word, Phrase, and Sentence Reading. 15 minutes.

Type 3. Reading of Paragraphs of Directions. 20 minutes.

The author recommends that the tests be given equal weight in making a composite score.

Let us suppose that the *Detroit Word Recognition Test* and the *Haggerty Reading Examination, Sigma 1*, have been given to pupils at the end of Grade 1H or at the beginning of Grade 2L.² According to criterion (1), the trait measured, these tests are judged to have equal significance, and each will receive a weight of 1. Before judging the tests according to criterion (2), reliability, we must estimate the *actual* working time. Since the tests have been given to pupils ready to enter the second grade, most pupils will be unable to attempt the entire *Haggerty Reading Examination* on account of the difficulty of the material at the upper end of the test.

¹ The reliability of most standard tests published up to the year 1927 is reported in Kelley, Truman L., *Interpretation of Educational Measurements* (World Book Co., 1927). Many authors of tests do, and all authors should, report reliability in the manual of directions or in an article appearing in a technical magazine.

² For example, 2L means 2 Low, that is, the Low Second Grade, or the first semester of the second grade. In some parts of the country the Low Second Grade is called 2A; in others 2B.

The average working time is about 12 minutes. According to criterion (2), therefore, the *Detroit Word Recognition Test*, for which the working time is 4 minutes, receives a weight of 1; and the *Haggerty Reading Examination*, for which the actual working time is about 12 minutes, receives a weight of 3. Keeping in mind both criteria, we assign the *Detroit Word Recognition Test* a weight of 1; the *Haggerty Reading Examination* a weight of 3.

Let us now suppose that the same tests have been used at the beginning of Grade 2H. For this grade, most teachers feel that the test measuring comprehension should have more weight than the one measuring word recognition. However, in considering only the trait measured, it is impossible to set up a rule for determining whether the weight should be 2 or 3 or more. A decision must be made in the light of the course of study, and other local conditions. According to criterion (1), the *Detroit Word Recognition Test* is judged to be worth a weight of 1; the *Haggerty Reading Examination*, a weight of 2. Before judging the tests according to criterion (2), we again estimate the actual working time, and find it to be 4 minutes for the *Detroit Word Recognition Test* and 16 minutes for the *Haggerty Reading Examination*. According to criterion (2), therefore, the *Detroit Word Recognition Test* is judged to be worth a weight of 1; the Haggerty test, a weight of 4. Keeping in mind both criteria, we assign the *Detroit Word Recognition Test* a weight of 1; the *Haggerty Reading Examination* a weight of 5.

In conclusion, the reader should note that tests should be given equal weights in case of doubt. Equal weights will seldom introduce any serious error.

Step 2. Obtain the Weighted G Score.—Let us assume that a pupil in Grade 2H has a Gr of 2.6 on the *Detroit Word Recognition Test* and a Gr of 3.0 on the *Haggerty Reading Examination*. The weights to be used are 1 and 5, respectively. The formula to be used in combining the Gr scores is as follows:

$$\text{Mean Gr} = \frac{w \text{ Gr}_1 + w \text{ Gr}_2}{\text{sum of the } w\text{'s}}$$

The definitions of the elements in the formula are as follows:

Mean Gr = average Gr obtained by combining two or more Gr's

w = weight assigned

Gr1 = G score on one reading test

Gr2 = G score on another reading test

sum of the w's = sum of the weights assigned

To illustrate in the case of a pupil in the 2H Grade whose Gr on the first test (*Detroit Word Recognition*) is 2.6 and whose Gr on the second test (*Haggerty Reading Examination*) is 3.0, the formula becomes

$$\text{Mean Gr} = \frac{1(\text{Gr1}) + 5(\text{Gr2})}{6}$$

Substituting

$$\text{Mean Gr} = \frac{1(2.6) + 5(3.0)}{6}$$

$$\text{Mean Gr} = \frac{2.6 + 15.0}{6} = 2.9$$

It is obvious that if the tests are given equal weight, the formula becomes

$$\text{Mean Gr} = \frac{1(\text{Gr1}) + 1(\text{Gr2})}{2}$$

$$\text{Mean Gr} = 2.8$$

This is simply averaging the two scores, a procedure with which every teacher is familiar.

In conclusion, it is recommended that fractional weights, such as 2.5, be avoided. Usually the data do not warrant such fine discriminations. If, however, it is desired to assign weights to three tests of 3, 2, and 1.5, respectively, it is better to use 6, 4, and 3. Time will thus be saved in making calculations.

The process just described may also be used in combining G scores on two or more intelligence tests, two or more arithmetic tests, and the like.

The G scores are now ready to be recorded on the Class Record Sheet.

The procedure is the same for combining age scores, although one uses either age scores or grade scores and not both. Grade scores are more convenient for grouping pupils.

CHAPTER X

HOW TO PREPARE CLASS RECORD SHEETS

Record sheets are usually provided by publishers for each of their tests. In order, however, to get a complete picture of every child and of every class, it is desirable to tabulate all pertinent data, such as age and G scores, on a single sheet, which may be termed a Class Record Sheet. This has been done for a sample class (see Table 4). In studying Table 4, the reader should bear in mind the fact that the sample school promotes pupils annually. Attention is also called to the fact that it is a small school; it has only one third grade, one fourth grade, and so on. The record sheets for Grades 1, 2, 4, 5, and 6, which are not reproduced here, are similar to that of Grade 3. (If the reader is interested in a school organized on a semester-promotion basis, he should read Chapter XII after reading this chapter.)

First take a sheet of ruled paper $8\frac{1}{2}$ " by 11" or $8\frac{1}{2}$ " by 13" and record at the top the desired identifying information, such as Grade, School, Date, Teacher's Name, Room Number, etc. The procedure in constructing a Class Record Sheet is as follows.

Step 1. Record Names or Numbers.—Arrange the names of the pupils by classes and grades. Within each class, record the names alphabetically, last names first. No names should be omitted, even though a pupil may have been absent for one or more tests. It is desirable that tests be administered to all absentees. In Table 4 numbers have been used instead of names to facilitate references and to avoid publicity for the children. Teachers should be especially careful not to discuss publicly the children's scores, particularly Gi.

Step 2. Compute and Record Grade Norms.—Column 2 of Table 4 contains the grade norms, i.e., G grade. The G table for a test is so constructed that the norms represent the average achievement of pupils of a given grade at a given time. The norm for any G score depends upon the time of year when the

tests are given. For any standard test, the G score norm for the beginning of the third grade is 3.0; for the beginning of the fourth grade, 4.0; and so on. Each additional month simply adds 0.1. For example, if the tests are given on December 5, or after three months of instruction, the third-grade norm will be 3.3. Similarly, the fourth-grade norm will be 4.3. If the tests are given on March 20, the third-grade norm will be 3.7. The G grade would, of course, be 3.2 instead of 3.7 if the class began the work of the third grade on February 1. For convenience,

TABLE 4

CLASS RECORD SHEET

FOR SCHOOL HAVING ANNUAL PROMOTION

Horace Mann School. Sept. 9. Miss L. Grade 3L.

(1) Pupil No.	(2) G grade	(3) G age	(4) Gi	(5) Gr	(6) Ga	(7) Gs	(8) Ge	(9) Gp	(10) Stat ^a	(11) Classification Conserv ^a
1	3.0	2.1	3.4	2.1	3.5	2.1	2.6	2.9	2H ^b	3L ^b
2	3.0	3.0	4.2	4.7	3.6	2.5	3.6	3.8	3H	3L
3	3.0	2.5	4.6	4.7	4.0	3.3	4.0	4.2	4L	3L
4	3.0	2.5	4.2	5.3	3.2	3.7	4.1	4.1	3H	3L
5	3.0	2.1	3.8	3.3	4.2	5.1	4.2	4.1	3H	3L
6	3.0	3.0	3.6	2.0	3.5	3.3	2.9	3.1	3L	3L
7	3.0	2.1	5.6	4.7	3.6	5.2	4.5	4.9	4H	4L
8	3.0	3.8	5.8	4.2	4.2	5.2	4.5	4.9	4H	4L
9	3.0	3.7	3.8	4.0	4.5	3.5	4.0	3.9	3H	3L
10	3.0	2.1	3.2	3.3	3.1	2.4	2.9	3.0	2H	3L
11	3.0	2.7	3.6	2.6	3.5	2.2	2.8	3.1	3L	3L
12	3.0	3.2	5.4	4.9	4.5	4.2	4.5	4.8	4H	4H
13	3.0	2.2	4.8	4.3	3.4	3.4	3.7	4.1	3H	3L
14	3.0	2.1	4.6	4.9	5.3	4.5	4.9	4.8	4H	4L
15	3.0	3.3	3.0	2.1	3.4	1.9	2.5	2.7	2H	3L
16	3.0	3.0	4.2	3.3	3.4	2.9	3.2	3.7	3H	3L
17	3.0	3.1	3.2	2.0	3.4	3.5	3.0	3.1	3L	3L
18	3.0	2.2	3.4	4.0	3.4	3.9	3.8	3.7	3H	3L
19	3.0	2.8	3.6	2.9	4.0	2.6	3.2	3.3	3L	3L
20	3.0	3.3	4.2	3.3	4.9	2.6	3.6	3.8	3H	3L
21	3.0	3.3	3.2	1.4	3.4	1.7	2.2	2.5	2L	3L
22	3.0	3.8	4.2	2.0	4.9	3.4	3.4	3.7	3H	3L
23	3.0	3.0	2.8	2.3	3.7	1.9	2.6	2.7	2H	3L
24	3.0	2.5	3.2	1.8	3.3	2.4	2.5	2.7	2H	3L
Total	72.0	65.4	95.6	80.1	91.9	77.4	83.1	87.5		
Mean	3.0	2.7	4.0	3.3	3.8	3.2	3.4	3.6		

^a Stat.—Statistical. Conserv.—Conservative^b L—Low. H—High For example, 3L means 3 Low, or the Low Third Grade, or the first semester of the third grade. In some parts of the country Low Third Grade is called 3A, in others 3B

the norms are counted to the *nearest* month, that is, the norm changes from 3.0 to 3.1 on September 16. The proper G grade may be read directly from Table 11.

The G grade tells how much to expect of each pupil in view of the grade he is in and the length of time he has been in it.

Both here and in subsequent chapters, the reader is advised to treat an eight-months school or any other number of months just as if it were a ten-months school. (For explanation of procedure in schools where the school year is shorter than ten months, see Section 6, "The School with a Short Term," in Chapter XIII.)

If age scores are being used instead of G scores, first determine G grade, then find it in Column 5, Table 21, and convert it into the corresponding age score in Column 1. Record in Column 2 of Table 4.

Step 3. Compute and Record Chronological Age.—Compute the chronological ages of the pupils, expressed in years and decimals of a year. In computing ages, it is necessary to compute them as of the date on which the tests are given. It is assumed in Table 4 that all tests are given on the same or approximately the same day. It is customary to drop any number of days less than half a month; half a month or more is regarded as a whole month.

Ages in years and months may be converted into years and decimals of a year by means of the conversion table on page 279.

Find the first pupil's age in Column 1, Table 21, and convert it into the corresponding G score. Thus an age of 10.4 converts to a G score of 4.6.

This is the pupil's *G age*, i.e., his age norm expressed in G scores. Record it in Column 3 of the Class Record Sheet (see Table 4). Proceed similarly for the other pupils.

The *G age* tells how much to expect of each pupil in all tests in view of his age.

Both here and elsewhere it should be remembered that G scores cannot be compared unless they are all for the same or approximately the same date. If the interval between the dates for any two G scores is a month or more, the G scores should be adjusted before making comparisons.

If age scores are being used instead of G scores do not convert a pupil's age into anything. His age in years and decimals

thereof is the age score norm and should be recorded in Column 3.

Step 4. Record Gi.—In the fourth column are the Gi's or grade scores made by the class on the intelligence test. Chapter VIII tells how to determine the Gi.

Those readers who are more familiar with age scores should note that the score which corresponds to Gi is mental age (MA), which is read from a table just as is Gi.

Step 5. Record Gr.—The fifth column of Table 4 contains the Gr's or grade scores in reading. In the sample class, Pupil 14 has a Gr of 4.9 which means that his achievement in reading is equivalent to that of the average fourth-grade child throughout the nation after nine months of instruction. It will be noted that the Gr's, in this 3L grade, range from 1.4 to 5.3.

Gr corresponds to reading age (RA), which, like Gr, is read from a table.

Step 6. Record Ga.—The sixth column of Table 4 contains the Ga's, or grade scores in arithmetic. They range in this particular 3L grade from 3.1 to 5.3.

Ga corresponds to arithmetic age (ArA), which is read from a table.

Step 7. Record Gs.—The seventh column of Table 4 contains the Gs's, or grade scores in spelling. The range in this 3L grade is from 1.7 to 5.2.

Gs corresponds to spelling age (SA), which is read from a table.

Step 8. Compute and Record Ge.—The eighth column of Table 4, contains the pupils' Ge's, that is, grade scores in education, or G scores according to all the educational tests combined. Ge is computed according to the formula

$$Ge = \frac{wGr + wGa + wGs}{\text{sum of the } w's}$$

where w signifies weight. In our sample class (Table 4), each test is given a weight of 1.

To illustrate in the case of Pupil 14,

$$Ge = \frac{1(4.9) + 1(5.3) + 1(4.5)}{3}$$

$$Ge = 4.9$$

In the case of Pupil 1,

$$Ge = \frac{1(2.1) + 1(3.5) + 1(2.1)}{3}$$

$$Ge = 2.6$$

The weights should be the same for all classes which take the same tests.

As noted above, in the sample class (Table 4) the educational tests were given equal weight. It might be argued that spelling (Gs) should not have as much weight as reading (Gr) since the reading test covers a portion of the work of the school that is probably of greater significance than that covered by the spelling test. For similar reasons it might be argued that Gr should have more weight than Ga in computing Ge. No rule can be given for determining what are the best weights to use. If the actual working times are strikingly different, this fact may be the basis of a decision. In the light of the tests used, the time allowances, and the local school policies, the examiner must decide the weights for himself.

If all the tests cover significant portions of the curriculum of the school, however, and if it is desired to complete the clerical work quickly, the examiner will follow the procedure of giving equal weight to each test. This has been done in Table 4.

If only two tests have been given, the formula is thereby made shorter. Similarly, if geography, history, or other tests have been given in addition to tests in reading, arithmetic, and spelling, the formula is expanded accordingly. Additional columns must be allowed on the Class Record Sheet to provide for recording the scores on these other tests.

The technique described above may seem to be impracticable if a pupil has been absent for one or more of the educational tests. If it is at all possible, absentees should be assembled within a few days after the date of the testing program. Pupils from grades which are given the same tests may be given the tests at the same time by the principal or by a teacher appointed for the purpose. There may, however, be blanks in the class record sheet on account of continued absence of pupils or for other reasons. If only one G score—for example, Gr—is missing, it is suggested that the teacher estimate the missing score. This estimate may be made as follows:

a. The teacher may average the G scores obtained on educational tests actually taken by the pupil. If Gs and Ga are available, these may be averaged in order to get an estimated Gr.

b. The teacher who has taught a pupil for two or three months may estimate his G score on the basis of tests which she has administered. Suppose that the Gr, or grade score in reading, of Pupil 1 of Table 4 were missing. The teacher knows that the reading ability of this pupil is about the same as that of Pupils 6 and 15. Their Gr's are 2.0 and 2.1. Averaging these G scores, she would obtain an estimated Gr of 2.1 for Pupil 1.

All such estimated scores should be circled with red ink. Ge's which include estimated G scores should also be circled.

The range of Ge's for this 3L grade is from 2.2 to 4.9. Ge corresponds both in its *general* interpretation and in method of computation to educational age (EA) in the age scale system.

CHAPTER XI

HOW TO CLASSIFY PUPILS

Step 9. Select the Basis for Grouping Pupils.—Many possible bases have been proposed. These are described and discussed, assuming first a school which has one class in each of Grades 1 to 7 or 1 to 9.

Basis 1. Grouping by chronological age.—An increasing number of educators, though still very much in the minority, propose to group pupils, even in senior high school, by chronological age and promote each year 100 per cent of the pupils. They claim that this method is ultra-simple, avoids embarrassment to teachers, pupils, and parents, compels teachers to individualize teaching and diversify the learnings, and encourages the use of an activity program which alone can utilize in a class project all kinds of talent and all levels of ability.

Basis 2. Grouping by achievement or achievement and intelligence status.—Other educators claim that, whether a school is an activity school or a traditional one it is better to put together pupils of approximately equal educational status or a combination of educational and intelligence status. They believe that homogeneous groups will make more satisfactory progress, due to the fact that the teacher can teach such a group almost as one pupil. The needs of all pupils are then closely similar. The work can be more exactly adapted to all. It saves the wear and tear on the teacher of continually shifting adjustment from one grade of ability to another. Franzen has described the instruction of teachers in non-homogeneous groups thus, "they mystify the lower quarter and bore the upper quarter."

There has been some controversy, especially in the elementary school, over the question: Should pupils be grouped by educational age (or Ge) or by mental age (or Gi).

Educational age when determined by a proper team of educational tests is probably superior to mental age for realizing the first objective of *bringing together pupils of equal educational*

status. Educational age is superior to mental age for this purpose because it and it alone reveals directly what pupils are of equal status educationally. Educational age measures this directly. Mental age measures educational status only indirectly. There is a close relation between mental age and true educational status, but there are many forces operating to prevent this correlation from being perfect. A pupil's educational status is a resultant not only of his mental age but also of his health, attendance, attitude toward school work, industry, etc. Educational age takes into account both mental age and all these other factors which condition the quality of school work. Mental age, as usually tested, reveals the effect of these other factors but to a less extent.

Again, educational age is superior because it prevents the pupil from skipping valuable portions of the curriculum. If the curriculum has been properly constructed most of what is ahead is not likely to be so valuable as an equal amount of what is behind.

Finally, educational age is superior because it prevents the skipping of prerequisite portions of ability hierarchies. Work in the elementary school is of a rather hierarchical nature. Even geography and history have certain prerequisites only a short distance below them. This point should not be stressed too much because gifted pupils have a phenomenal capacity to fill up really vital gaps. But educational age, particularly when it rests upon educational tests for the more continuous subjects, does guarantee that the pupil will not be handicapped by large gaps in his abilities.

Franzen, trying to prove the superiority of mental age, demonstrated, in the case of pupils whose educational age is markedly below mental age, that by specially promoting them and by otherwise applying educational pressure the educational age could be made to approximate the mental age within one year. It would be interesting to learn whether this progress could not have been secured just as well, if not better, by keeping them at all times in the grade or grades closest to their educational age and applying the pressure there.

Mental age is, however, superior to educational age for classifying pupils in the primary grades, and possibly for high school and college freshmen also, though some schools follow

the practice of determining classification on the basis of educational tests of the progress made during the first week or weeks of school. For the lower primary grades, a reading readiness test may be superior to either an intelligence or educational test.

Since neither educational age nor mental age is ever measured adequately and since there is a close correspondence between them, it is probably better to combine them for purposes of grouping. The author advises this.

Basis 3. Grouping by chronological age and achievement.—Most educators advise a small school, even an activity school, to use a compromise between *Basis 1* and *Basis 2*. For this reason Book Three provides the appropriate techniques more fully for it than for any other basis.

Since *Basis 3* finds a large place for mental and educational measurement, whether determined by standard tests or teacher's examinations and opinions, it is well to consider some objections to this plan.

1. *Young pupils are forced to compete with the mentally more mature.*—This is a relic of the old notion that all pupils are born equal and that subsequent mental age keeps pace with chronological age. In general this objection represents a misplaced sympathy. Every investigation shows that it is a rule for the young pupils to be leading their classes and for the older pupils to be struggling to keep up.

Witty and Wilkins¹ summarized the investigations bearing on grade skipping and concluded that most studies "show clearly that acceleration is associated with desirable adjustment in *all* types of development for which data have been assembled." They strongly urge that grade skipping be more generally practiced especially in small schools, and that supplementary athletic and social grouping be provided, if needed, partly by playground adjustment, partly by inter-grade social provisions, and partly by skipping more bright, young children to keep the previously skipped ones company.

2. *Young pupils have difficulty in making social adjustments.*—It would be truer to say that older pupils have difficulty in

¹ Witty, Paul A. and Wilkins, Laroy W., "The Status of Acceleration or Grade Skipping As an Administrative Practice," *Educational Administration and Supervision*, May, 1933.

adjusting to the younger ones. There is an undoubted tendency for older pupils to dislike the presence of a much younger pupil. How serious this jealousy is needs to be investigated.

But what will the gifted pupils do when they reach the high school while still very young? One suggestion is that they delay their arrival at the high school by taking a wider educational swath. If, however, the curriculum has been properly constructed this means that the gifted pupil will be spending almost half his time upon material of relatively small value. The only satisfactory solution is to provide a path for the geniuses which leads from the first grade through the university, so that the genius pupil may be with his kind throughout his entire educational career. If he graduates from the university while still too young he may be employed in national research or on other large social enterprises until he is judged sufficiently mature physically to take his place in the general social group.

The only visible solution for the small school is to promote the young gifted pupil just as often as the older pupils will permit without making his life miserable. Another solution is to abolish the school which is too small to make adequate provision for individual differences among its pupils and to substitute the consolidated school in its place.

3. *Causes vital gaps in pupil's education.*—Classification by educational age meets this objection provided the testing has been thorough. To receive a high educational score shows that these gaps have somehow been filled.

It is difficult to believe that this is the real objection. It can be demonstrated that older pupils have phenomenally large gaps in prerequisite abilities. But this does not seem to produce any particular concern. Worry comes only when the young pupil is involved. Educators find it almost as difficult as laymen to prevent themselves from thinking in terms of such irrelevant surface factors as chronological age, physical size, and brute muscles. We think of children as we would of elephants or dinosaurs. Considering how much of our lives are regulated by chronological age, this is not surprising. We are born at zero years of age, compelled to begin school at six, permitted to leave school at fourteen, allowed to marry at sixteen, entitled to vote at twenty-one, and are given an average salary at that age where a long life of usefulness is passing into decline.

Squeezed in between a chronological end and a chronological beginning the passage through school naturally becomes a chronological procession.

4. *Disregards health.*—Some picture the gifted child as a frail, forced, hot-house flower. Terman, after a careful study of many gifted pupils, concluded that they were no more frail than ordinary children. He found that some were frail and some robust. Consequently if there is any reason to suppose that health will be sufficiently improved by giving the pupil intellectually easy tasks, health should certainly be considered.

There is, however, a fear abroad that a pupil's mind may, like Jefferson's Constitution, be stretched until it cracks. In his Columbia University Master's thesis Franzen describes a ten-year-old pupil in Grade V with an I.Q. of 178. This genius distinguished between poverty and misery, thus: "Poverty is the lack of things we need, misery is the lack of things we want." He defined a nerve as the "conduction unit of sensation" and explained correctly what he meant thereby. It was discovered that he had read all the textbooks of the grades ahead of him. His two able parents after a consultation with the family physician refused permission for him to be promoted from the grade where he was bored almost to extinction because it might *strain his mind!*

5. *Emphasizes the intellectual to the exclusion of character traits.*—This is another way of saying that pupils are classified by their abilities and not by their purposes. It may be that a pupil can be taught desirable purposes in one grade as easily as in another. It is certain that purposes do not fall into such close hierarchies as do abilities. Furthermore, it is possible that most pupils who are promoted for their intellectual achievements would likewise be promoted for their composite character status. Terman¹ studied the extent to which intellectually gifted pupils possessed the following intellectual and personal traits: sense of humor, power to give sustained attention, persistence, initiative, accuracy, will power, conscientiousness, social adaptability, leadership, personal appearance, cheerfulness, coöperation, physical self-control, industry, courage, dependability, self-expression through speech, intellectual mod-

¹ Terman, L. M., *The Intelligence of School Children*, p. 58; Houghton Mifflin Co., New York City, 1919.

esty, obedience, popularity among fellows, evenness of temper, emotional self-control, unselfishness, and speed. Any reader would not complain of any lack if he possessed intelligence plus this galaxy of traits. Terman found that all these traits correlated positively with intelligence, that is to say, with ability primarily. The first trait, sense of humor, has, in the case of gifted children, a correlation of .58. The last trait, speed, correlates .28. The others gradually vary between these extremes in the order named. Terman claims that he can roughly predict I.Q. from an average of these 24 traits. Subsequent studies tend to confirm his claim.

The most common explanation given by teachers for the failure of certain specially promoted pupils to do satisfactory work is that they do not try. It is generally admitted that they could do the work of the grade if they only would. These remarks by teachers suggest two questions. (1) Since tests reveal that these pupils have actually mastered, somewhere, somehow, large segments of the curriculum, is it not possible that they are mastering the material of the new grade with such unobtrusive ease as to deceive even the keenest observer? (2) If the teachers are correct may not the pupil's lack of industry be due to improper habits formed by previous improper classification where industry was not required?

Basis 4. Grouping by attainment and sectioning by chronological age.—A large school can be organized on any one of the preceding bases or on another not available to a small school, namely one which classifies into grades on the basis of measured attainment and into sections within the grade on the basis of chronological age.

Violent controversy has raged over this problem of XYZ or I.Q. or age or homogeneous grouping into sections within a grade.

The opponents claim that XYZ grouping leaves the teacher facing individual differences in specific abilities almost as wide as before the grouping occurred. They hold that homogeneity in general ability does not prevent marked heterogeneity in specific abilities, and that it is specific abilities which are taught. They quote studies which do show that much heterogeneity remains, though not as much as previously. The proponents answer that these studies have generally failed to take into

account the fact that a group's variability is increased beyond the true variability by an unreliable test—and all tests are somewhat unreliable. Generally, the true variability is one-tenth to one-fifth less than the apparent variability.

Worse still, insist the proponents, the investigators have failed to take cognizance of a much earlier and far more discerning dissertation by Franzen, who gave a fairly convincing demonstration that the variability found in specific subjects after homogeneous grouping by general mental ability is not rooted in the inherent natures of pupils but is largely due to the non-homogeneous grouping during past school years. Except for minor inherent idiosyncracies he brought each pupil's specific ability on a level with his general mental ability after two years of greater homogeneity in grouping. Finally the proponents point out that the heterogeneity found by Akridge and others is magnified because the sectioning investigated was based mainly on teacher's subjective estimate rather than on objective tests.

The opponents hold that XYZ grouping is offensive to principals, teachers, and especially parents. Sauvain pretty thoroughly killed the notion that homogeneous grouping is offensive to those in school and out by discovering that parents, pupils, principals, and teachers strikingly preferred it after having had experience with both it and heterogeneous grouping. He had the results of his study carefully sorted and tabulated in various ways. Critical ratios were computed to determine the importance of differences. His findings ¹ follow:

FINDINGS

General Findings:

1. The large returns of teachers, parents, principals, and other school officials from the 16 cities participating indicate that the problem is one of considerable interest to them.

2. The interest shown seems surprising since the subject has received so little investigation in the past and many cities are unwilling to have the subject investigated now.

Several of the cities asked to coöperate have been moving away from the use of grouping or have abandoned it altogether. The study has no

¹ Sauvain, Walter H., *A Study of Opinion Regarding Homogeneous or Ability Grouping*, Bureau of Publications, Teachers College, Columbia University, New York City.

evidence gathered on the extent of the movement in the opposite direction.

3. The techniques involved in collecting data have resulted in unusually high returns as questionnaire studies go.

Opinions of Parents:

1. On the whole, parents seem favorable to the use of grouping where it is employed.

This is especially true of those with children in bright groups.

There is more parent opposition than would be indicated by principals' estimates of parent complaints.

2. Many more parents say they know in which ability sections their children are located than do correctly state the sections of their children.

More than twice as many parents of children in slow groups state the sections of their children incorrectly as state them correctly. Parents knowing and admitting that their children are in slow groups are more opposed to grouping than parents of other children in slow groups. Parents knowing and stating that their children are in bright groups are far more in favor of grouping than other parents having children in bright groups.

3. On the whole, where grouping is used, parents believe that children are at least as happy, do better work in school, and are correctly sectioned according to ability.

Many parents not objecting to the school's placement have urged their children to get into higher ability groups. Over four-fifths of all parents indicate that they believe their children know in which ability sections they are located.

Opinions of Teachers:

1. Teachers seem to favor ability grouping somewhat more than do the parents.

2. Teachers preferences as to ability sections vary widely, although slow groups are the least popular.

Bright and average groups are considered about equally desirable. About a fifth of the teachers in charge of slow ability sections would never teach there if given a choice in the matter.

3. Teachers in most communities report themselves quite well satisfied with grouping as it is there practiced.

Less than 5% state they would abandon grouping. Less than half believe that serious changes are needed in grouping. Of teachers with experience under both ability grouping and heterogeneous grouping, over 90% say they prefer the use of "ability grouping" rather than "no grouping other than grades."

Opinions of Principals and School Officials:

1. Principals and other school officials are fully as well pleased with the ability grouping as the teachers.

Factors Related to Responses about Grouping:

1. The section in which the child is located is an important factor. Teachers and parents of children in bright sections are more in favor of grouping than teachers and parents of children in other sections.

2. The basis on which grouping is done has some relationship to the responses.

Where the I.Q. is weighted heavily in doing the sectioning, teachers are not so sure that desirable social attitudes result. Parents are definitely more favorable to grouping in schools where the I.Q. is weighted 50% or more in doing the grouping.

3. The educational philosophies indicated by teachers do not seem to bear important relationships to their responses.

4. The apparently greater approval by parents who otherwise indicate progressive philosophies is probably due to the fact that their children are largely in bright sections.

5. Where the curriculum has been definitely adapted to meet the needs of ability groups, teachers are more in favor of grouping.

6. Adaptations of the curriculum to meet the needs of ability groups do not affect parent opinion favorably.

7. The type of community in which the school is located bears some relation to the responses of parents and teachers.

Teachers in low-class communities seem surer that desirable social attitudes accompany grouping and are more inclined to avoid teaching slow sections. Parents in high-class communities are more in favor of grouping than those in less favored areas.

8. Teachers show a slightly greater preference for grouping in schools where opportunities are not restricted for slow groups by decreasing enrichment and increasing time spent on minimum essentials.

The school's policy on this matter does not seem to be related to the responses of parents.

9. The sex of the parent answering the blank or of the child concerned seems to play relatively little part in the answers of parents cooperating.

10. The grades in which the children are located bears some relation to the answers of both teachers and parents.

Teachers of lower grades find grouping more to their liking than do those of upper grades. Parents of upper grade children seem surer that grouping leads to better school work and yet exert less pressure on their children to get them into higher ability groups.

11. Teachers with experience only under ability grouping like it better than those with experience under both ability grouping and heterogeneous grouping.

12. Few teachers admit telling their sections what ability levels these represent.

Probably their telling the sections does not seriously affect parent opinion.

The opponents insist that sectioning is a violation of the principle that school should duplicate the whole of life. They contend that adults do not stay in *one* group but move freely from group to group according to their *interests*. The proponents reply that they favor reproducing in the school only the desirable aspects of the life of society in general and not all of it, that the pupils are in their class sections only four hours out of every twenty-four, that during these hours they have, in the better schools, a considerable amount of association with pupils in other sections and persons outside of school, that adults themselves would probably form themselves into similar sections if they were learning such abilities as the pupils are learning.

The opponents, notably McGaughy,¹ hold that Thorndike's dictum that good things tend to go together has insidiously lured us into an acceptance of sectioning. McGaughy contends that the correlation between good things, say intelligence and reading ability, is much too low to justify sectioning on the basis of intelligence. He finds that the correlation between one intelligence test and one reading test is far from perfect being only about .75, and that it is much lower for other abilities. The proponents say that he fails to recognize that a correlation of much less than perfect, i.e., 1.0, is adequate for forming reasonably satisfactory sections, since the majority of high intelligence pupils would be safely above the line where the sections are separated. To demand a standard of 1.0 is to assume that there are to be as many sections as there are pupils, and that absolute perfection of sectioning is absolutely necessary. They say further that he ignores the fact that Franzen has proved that the correlation is much higher when there had previously been sectioning on the basis of intelligence tests. Again, they say that he forgets that when the correlation of .75 is corrected for sheer error in the measurements the correlation coefficient approaches unity, and that any suitable system of cumulative records enables one to realize the benefits of such a correction. Also they say that he is battling a straw man with all the courage of a Don Quixote since it is not the common practice, nor the approved practice, to base decisions about classification

¹ McGaughy, J. R., *An Evaluation of the Elementary School*, The Bobbs-Merrill Company, Indianapolis, 1937.

on a single intelligence test, but rather to make decisions either in the light of cumulative records or a battery of contemporary tests or both, thus greatly reducing the error. Finally, the proponents claim to have discovered that there is a higher correlation than he believes exists between intelligence and, say, handwriting at that stage of a pupil's progress where good handwriting is stressed and highly approved. When it ceases to be of major importance in instruction the correlation drops. But this does no particular harm, since it is of no particular consequence if a pupil in the highest section, say, exhibits only average or below average penmanship. And this tends to be equally true of many other traits such as appreciation of art and music, although these may be very important in the life of the child and the philosophy of the school.

The opponents of XYZ grouping and the press generally have widely publicized the abandonment of homogeneous grouping by the Horace Mann School. But, answer the proponents, the school had homogeneous grouping after abandoning it. When practically all the pupils in the school have I.Q.'s over 100, homogeneous grouping is a substantial reality already. The public school, having half its enrollment below 100 I.Q., faces a very different situation.

The opponents claim that XYZ grouping, reducing the heterogeneity somewhat, will cause teachers to neglect to adapt the curriculum to each and every individual as a separate person with his own idiosyncracies, unique problems, and special talents. The proponents hold that properly conceived XYZ grouping will make simpler the teacher's task of making adaptations to individual differences.

The opponents of XYZ grouping claim that such segregation is undemocratic. The proponents claim that life in our democracy exhibits all kinds of grouping, many of which are on the basis of ability. They insist that equal or optimum opportunity and not identity of treatment is the essential principle of democracy. They favor a plan of grouping which permits the ready shifting of pupils from group to group when fuller knowledge indicates the wisdom of doing so. Teachers agree that when the grouping is properly administered and teachers sympathetically stress the special talents of the slow pupils, pupils do not feel stigmatized. Furthermore, the proponents advocate

providing for all sorts of other transient or permanent special groupings which bring together all types and ages of pupils, such as dramatic clubs, music clubs, athletic teams, whole school projects, inter-class activities, pupil government, *et cetera*.

The opponents claim that XYZ grouping fails to give due respect to pupil personality. The proponents claim that placing pupils with their kind shows greater concern for their personality than placing them in mixed groups where the dull have their dullness continually emphasized. They also point out that a census of pupils' opinions shows that pupils prefer homogeneous grouping, and that an investigation shows a reduction in the number of disciplinary problems following the adoption of such a grouping.

The opponents claim that XYZ grouping will develop undesirable feelings of superiority among the gifted and of inferiority among the dull pupils. The opinion of teachers who have had experience with XYZ grouping was canvassed relevant to this point. One study found the teachers about equally divided, but the others find that teachers' opinion is generally favorable to XYZ grouping.

The opponents claim that XYZ grouping, based as it usually is on intelligence or achievement tests, fails to take into account the "whole child"—his attitudes, appreciations, purposes, and the like. The proponents reply that there are excellent reasons for grouping on the basis of skills and abilities since these limit a pupil's capacity to profit by instruction, whereas there is little reason to believe that character education has a hierarchy comparable to the skills or can be provided for much better by one grouping than another, and that what little evidence there is favors XYZ grouping.

The opponents claim that XYZ grouping requires, if it is to be specially helpful, a teacher specialization that is difficult to secure in practice. The proponents answer that considerable progress has been made in developing specialized teachers for dull classes and that the general practice of XYZ grouping will intensify the effort to develop specially trained teachers.

The opponents claim that XYZ grouping has caused no greater growth on the part of pupils. The proponents concede that a summary of *all* investigations leads to no sure conclusion

but that an analysis of the more carefully controlled experiments reveals a clear but small gain in favor of XYZ grouping, as, for example, the investigation by Barthelmeß. A careful investigation by Hollingworth of equivalent segregated and unsegregated bright pupils failed to show greater gain on test scores for the segregated pupils, but there had been curriculum enrichment without loss in test traits. Also there is usually a reduction in the number of failures following the introduction of homogeneous grouping. This may possibly be evidence of greater growth. The proponents hold that a little gain is encouraging since the advantage in favor of segregated groups will increase as tests become more comprehensive and curriculum adaptations are made.

But after all, how pupils are taught and not how they are grouped is the vital matter, so it is hardly worth laboring the argument. Those who care to consider the question further will find a comprehensive treatment of it in the *Thirty-Fifth Year Book, Part I*, prepared by the National Society for the Study of Education and published by the Public School Publishing Company, Bloomington, Illinois.

Basis 5. Semi-annual classification.—The semi-annual system of classification has been widely adopted in the hopes that it would correct the faults of the annual system. Lindsay¹ after an admirable analysis concluded that the arguments were about equally balanced. He asked which system could be most depended on to:

1. Conform with community custom.
2. Conform in general with practice; provide better placement of transfer pupils.
3. Accommodate more effectively temporarily absent pupils.
4. Enhance entrance by multiple entrance dates with less social age variation.
5. Distribute enrollment load over school year.
6. Relieve overcrowding in primary grades.
7. Advance new pupil group to better teachers to replenish depleted quotas.
8. Decrease congregation of truants and indifferent pupils in middle grades.
9. Provide flexibility which is claimed for homogeneous grouping.
10. Make possible homogeneous grouping.

¹ Lindsay, J. Armour, *Annual and Semi-Annual Promotion*, Bureau of Publications, Teachers College, Columbia University, New York City, 1933.

11. Constrict pupil-ability-social-range when ability grouping is lacking.
12. Obviate loss of time in reorganization at mid-year.
13. Provide opportunity for complete reorganization in summer.
14. Allow a greater degree of ease in administration.
15. Help to hold pupils in school longer.
16. Permit a better articulation of elementary and high schools.
17. Afford a better articulation of school and occupational life.
18. Afford flexible school organization with frequent adjustments possible.
19. Afford administrative relief in needed shifting of pupils and teachers.
20. Provide ease in trial promotion and acceleration of superior pupils.
21. Result in more frequent staff judgment and evaluation of pupil achievement.
22. Release pressure on the slower pupil.
23. Reduce amount of retardation.
24. Afford suitable make-up work interval and facile union with summer school.
25. Avoid pupil discouragement and relaxation of effort with failure imminent.
26. Cause less parental objection when pupils are retained in grade for term.
27. Provide satisfactory age-grade placement.
28. Effect increased effort from real immediate goals and frequent accounting.
29. Call for definite curriculum fitted to pupil with short interest units.
30. Permit broad scope of instructional materials and use of larger learning units.
31. Offer opportunity for emphasis on "child" instead of on a set course of study.
32. Permit teacher to spend more time on slow and less competent pupils.
33. With fewer grade-levels in room, aid in better diagnostic-remedial work.
34. Aid in limiting grade-levels per teacher with individual work enhanced.
35. Bring variety of work which is inspiration to pupils and teachers.
36. With multi-grade levels in room, permit lower grades to learn much from others.
37. With multi-grade levels in room, permit pupil to develop habit of focusing on work.
38. Permit review of recent work with more effective results after failure.
39. Allow teacher to become a specialist in work of term.
40. Permit more teacher contacts and shorter period with unsuitable teacher.

41. Permit teacher to have pupils throughout the whole year.
42. Provide a situation where parents are more apt to know pupil's teacher.
43. Decrease school operation cost due to less and lower cost of work repeated.
44. Require a smaller teaching force.
45. Avoid small mid-year classes.
46. Make economies possible in the operation of the school plant.
47. Reduce school expenses.

In the best treatment of this subject the author has seen, Lindsay lists and discusses many other modifications of the traditional grade pattern, some to replace it, some to modify it, and some to supplement either it or another basic plan. Among these modifications are:

Basis 6. The all-year-plan.—This was tried in certain schools, located in undesirable districts of Newark, N. J., to keep pupils off the streets and out of gangs and speed up their progress through the grades. The pupils did enter high school earlier and achievement tests showed that they had made greater progress than pupils in ten-months schools in the same city when matched for age, intelligence, background, etc.

Basis 7. The review term.—This plan involves the provision of a review term at intervals in the grades. The better pupils skip it. The others use it to overcome shortages.

Basis 8. Promotion by subject.—This is familiar to all.

Promotion by subjects, though almost universal in practice in the high schools, is so inimical to an adequate guidance program and so seriously hampers a proper educational program for the pupil that it is destined to disappear or be seriously modified, when the high schools complete their emancipation from the domination of subject-centered colleges.

One type of modification is illustrated in the Horace Mann School, Teachers College, where Switzer, as the regular teacher of the class, merged many subjects into a core curriculum and utilized special teachers as they were needed. Lesson Unit No. 3 (consult Chapter XVII) tells how she and Reeves utilized, relevantly, foods, clothing, science, composition, arithmetic, reading, literature, and psychology in a single dynamic unit. At other times she merged subjects into a core curriculum by means of less dynamic units such as the study of Egypt. Lesson Unit 99 illustrates the less dynamic type of merging.

is organized into integrated units and not into subjects and grades. Each pupil moves from unit to unit as he is able.

Basis 12. The coöperative plan.—This scheme is called the coöperative plan because, say, five teachers, specialists in five areas of the curriculum, coöperatively plan for and work with about 200 pupils, integrating their work around large units. The pupils move in groups of about forty from teacher to teacher to consider aspects of each unit.

Basis 13. The platoon plan.—According to this plan, pupils were not only grouped into grades on the basis of achievement and into fast and slow sections but also into platoons. The first platoon arrived early and departed early. The second platoon arrived later and departed later. Two classes occupied the same classroom except that while one class was in it, the other class was in the shop, in the auditorium, or on the playground. Thus one school building was made to serve two elementary schools.

The Gary plan likewise provided for fuller plant utilization by having an elementary and high school in the same building.

Basis 14. Miscellaneous methods.—It will be sufficient to merely list some of these: (a) trial promotions, (b) outside aid in school or home of pupil, (c) ungraded classes designed to correct maladjusted pupils and return them to their normal groups, (d) classes for the permanently atypical, mentally or physically, (e) over-age classes, (f) non-English speaking classes, particularly at the beginning of school, (g) summer terms, and (h) special sessions for parents engaged in certain seasonal occupations. A treatment of special classes will be found in *Review of Educational Research*, April, 1936, and October, 1937.

Basis 15. The self-starting, self-grouping plan.—After considering these multitudinous methods, the author proposes that someone experiment with an ever-emerging, ever-fluid, pupil self-grouping plan. Any imposed plan of grouping tends to assume an imposed curriculum. Why not reverse the process? Why not, at the beginning of the year, regard the whole school from top to bottom as one large class with all teachers belonging potentially to every pupil? Then each pupil, under the suggestion, stimulation, and guidance of any or all teachers or his own inner urges, could begin one or more activities. Gradually groups would form around those activities of most vital interest

to the pupils. Sometimes beginning and graduating pupils would be found in the same group, as, for example, in some dramatization. Sometimes the whole school would be coöperating in the production of, say, a community festival. Often single individuals would be absorbed in something of interest to him alone. The teachers would weave in and out among these ever-emerging and ever-dissolving groups, helping the pupils when help would be educative, and scrupulously withholding help when it would not be educative, each teacher contributing to any pupil or group of pupils according to her talents and experiences, sensitively guiding pupils' activities in ever more significant directions. For pupil accounting and certain intimate guidance purposes, the pupils might be grouped into home rooms according to age, but for normal instructional purposes, the author proposes, subject to modification during trial, that we have a self-starting curriculum and such self-grouping as might emerge.

The remainder of Book Three is devoted to suggesting practical ways of doing better what most schools are doing now.

Step 10. Compute and Record Gp.—The next column in the table is headed Gp or grade score for placement. This is probably the most significant score in the whole table, since it combines all the others. To review briefly, the Gi gives us a measure of a pupil's general intelligence, that is, his level of learning ability. The Ge is a measure of achievement in reading, arithmetic, and other school traits, since it is computed by combining the grade scores on the tests in these traits. Since both Gi and Ge measure something which plays a part in determining success in school work, the best index to use in classifying or promoting pupils is a combination of the two. This combination is known as Gp or grade score for use in placement.

The formula for computing Gp is:

$$Gp = \frac{w Gi + w Ge}{\text{sum of the } w\text{'s}}$$

where w signifies the weight to be used.

Again we have the problem of determining what weights to use. The principles involved are the same as those discussed in connection with Gr in Chapter IX. The tests used in the sample class (Table 4) were the *Multi-Mental Intelligence Scale*, *Thorn-*

dike-McCall Reading Scale, Mixed Fundamentals in Arithmetic, and Morrison-McCall Spelling Scale.

In view of the significance of the abilities measured, for the purpose of classification, and in view of the fact that Ge in our sample school is a more reliable measure, representing as it does about twice as much working time as Gi, we have decided to give Gi a weight of 1 and Ge a weight of 2. The formula then becomes

$$G_p = \frac{1 G_i + 2 G_e}{3}$$

To illustrate, Pupil 1 has a Gi of 3.4 and a Ge of 2.6. Substituting, we have

$$G_p = \frac{1(3.4) + 2(2.6)}{3}$$

$$G_p = 2.9$$

In the primary grades, if an individual intelligence test score is available, it is recommended that Gi and Ge be given equal weight. In the case of group intelligence test scores in primary grades, it is probably better to give more weight to Ge. In the upper elementary grades, Ge will in most cases be given more weight than Gi.

A principal may desire to include Gt, or grade score according to teacher's estimate, in the formula for computing Gp. See Step 4, Chapter XII.

In the age scale system, promotion age (PrA) corresponds to Gp. The computation is similar.

Step 11. Compute Total and Mean Scores.—Having computed the various G scores of all the pupils in the class, the next step is the computation of the total and the mean class scores. It will be observed that the totals are placed at the bottom of the several columns in Table 4. This total score is obtained by adding the individual pupil scores and norms.

Below the total scores we find the means. The mean, or average, is calculated by dividing the total score in each case by the number of pupils.

If any G scores are missing only those pupils' records that are complete should be used.

Step 12. Determine Which Classification Table to Use.—The reader will observe that three classification tables are provided

in Tables 7, 8, and 9. Table 7 is intended for use with a school which accomplishes 0.9 of a standard grade's work per year. In other words, such a school covers only minimum essentials. Table 8 is for a school which accomplishes one standard grade's work in one year. Most schools will use this table. Table 9 is for a school which accomplishes 1.1 of a standard grade's work in one year. This table will be used by schools in which the intelligence and achievement of the pupils are far above the average. The procedure for determining the amount of accomplishment within a given school is outlined in the following paragraphs. It will be observed that the amount of accomplishment is not necessarily proportional to the length of the school year.

The next step in completing the Class Record Sheet is to determine which classification table to use. This can be done very easily. The steps in the calculation are as follows: First, record the mean Gp scores for all grades. Below them record the respective norms. The scores and norms in the case of our sample school are shown in Table 5.

TABLE 5
CALCULATION OF DIFFERENCES BETWEEN GP'S AND NORMS

	GRADE			
	3L	4L	5L	6L
Mean Gp.....	3.6	4.5	5.7	7.1
Norm Gp.....	3.0	4.0	5.0	6.0
Differences.	0.6	0.5	0.7	1.1
Total of Differences	2.9			
Mean Difference.....	0.7			

Then subtract, algebraically, each norm Gp from the corresponding mean Gp, and record the differences with the proper signs. Total these differences algebraically and record the result opposite Total of Differences. (In Table 5, this figure is obtained by adding 0.6, 0.5, 0.7, and 1.1.) Compute the mean difference

by dividing the total difference by the number of different scores. In Table 5, 2.9 is divided by 4. The mean difference, 0.7, is interpreted thus: This school averages 0.7 Gp, or seven months above the norm in these four grades. In some cases, the mean difference will be negative. This will signify that the school averages on the whole below the Gp norm.

Table 6 indicates which classification table to use:

TABLE 6
SELECTION OF CLASSIFICATION TABLE

IF THE MEAN DIFFERENCE IS		USE
Below	-0.5	0.9 Classification Table
Between	-0.5 and +0.5. . . .	1.0 Classification Table
Above	+0.5	1.1 Classification Table

In our sample school, the mean difference (0.7) is above +0.5; hence we shall use the 1.1 Classification Table (Table 9).

Step 13. Compute and Record Statistical Classification.—By statistical classification is meant the proposed grade placement of pupils, as determined by the Classification Tables (Tables 7, 8, and 9). Statistical classification is based on Gp. The Classification Tables provide a convenient instrument for determining the grade in which a pupil belongs. The tables are so constructed that they may be used at any time in the year that tests are given. The term “statistical” is used to distinguish this classification, which is based on the available statistical data, from a conservative or actual classification, which may be influenced by other considerations. The significance of these terms will become clear as the work proceeds.

We are now ready to operate the “Table for a school which attempts to do 1.1 standard grades per year, showing the automatic classification of pupils into grades on the basis of any G (Grade) Score.” This is Table 9.

It will be observed that the first column of the table contains the G scores. The other columns are headed 0, 1, 2, 3, 4, 5, and refer to the number of months the class has been in either the first half (Low division), or the second half (High division) of the grade. In the case of our sample class, Table 4, the classes of this school have been in their respective grades less than fifteen calendar days; therefore all readings will be made in

TABLE 7

TABLE FOR A SCHOOL WHICH ATTEMPTS TO DO 0.9 STANDARD GRADE PER YEAR, SHOWING THE AUTOMATIC CLASSIFICATION OF PUPILS INTO GRADES ON THE BASIS OF ANY G (GRADE) SCORE¹

G SCORE	NUMBER OF MONTHS CLASS HAS BEEN IN HALF OF GRADE IT IS NOW IN					
G	0	1	2	3	4	5
0.0	KL	KL	KL	KL	KL	KL
0.1	KL	KL	KL	KL	KL	KL
0.2	KL	KL	KL	KL	KL	KL
0.3	KH	KL	KL	KL	KL	KL
0.4	KH	KH	KL	KL	KL	KL
0.5	KH	KH	KH	KL	KL	KL
0.6	KH	KH	KH	KH	KL	KL
0.7	1L	KH	KH	KH	KH	KL
0.8	1L	1L	KH	KH	KH	KH
0.9	1L	1L	1L	KH	KH	KH
1.0	1L	1L	1L	1L	KH	KH
1.1	1H	1L	1L	1L	1L	KH
1.2	1H	1H	1L	1L	1L	1L
1.3	1H	1H	1H	1L	1L	1L
1.4	1H	1H	1H	1H	1L	1L
1.5	1H	1H	1H	1H	1H	1L
1.6	2L	1H	1H	1H	1H	1H
1.7	2L	2L	1H	1H	1H	1H
1.8	2L	2L	2L	1H	1H	1H
1.9	2L	2L	2L	2L	1H	1H
2.0	2H	2L	2L	2L	2L	1H
2.1	2H	2H	2L	2L	2L	2L
2.2	2H	2H	2H	2L	2L	2L
2.3	2H	2H	2H	2H	2L	2L
2.4	2H	2H	2H	2H	2H	2L
2.5	3L	2H	2H	2H	2H	2H
2.6	3L	3L	2H	2H	2H	2H
2.7	3L	3L	3L	2H	2H	2H
2.8	3L	3L	3L	3L	2H	2H
2.9	3H	3L	3L	3L	3L	2H
3.0	3H	3H	3L	3L	3L	3L
3.1	3H	3H	3H	3L	3L	3L
3.2	3H	3H	3H	3H	3L	3L
3.3	3H	3H	3H	3H	3H	3L
3.4	4L	3H	3H	3H	3H	3H
3.5	4I	4L	3H	3H	3H	3H
3.6	4L	4L	4L	3H	3H	3H
3.7	4L	4L	4L	4L	3H	3H
3.8	4H	4L	4L	4L	4L	3H
3.9	4H	4H	4L	4L	4L	4L
4.0	4H	4H	4H	4L	4L	4L
4.1	4H	4H	4H	4H	4L	4L
4.2	4H	4H	4H	4H	4H	4L

¹ Table prepared with the assistance of Grace Moffatt.

TABLE 7 (Continued)

G SCORE	NUMBER OF MONTHS CLASS HAS BEEN IN HALF OF GRADE IT IS NOW IN					
G	0	1	2	3	4	5
4.3	5L	4H	4H	4H	4H	4H
4.4	5L	5L	4H	4H	4H	4H
4.5	5L	5L	5L	4H	4H	4H
4.6	5L	5L	5L	5L	4H	4H
4.7	5H	5L	5L	5L	5L	4H
4.8	5H	5H	5L	5L	5L	5L
4.9	5H	5H	5H	5L	5L	5L
5.0	5H	5H	5H	5H	5L	5L
5.1	5H	5H	5H	5H	5H	5L
5.2	6L	5H	5H	5H	5H	5H
5.3	6L	6L	5H	5H	5H	5H
5.4	6L	6L	6L	5H	5H	5H
5.5	6L	6L	6L	6L	5H	5H
5.6	6H	6L	6L	6L	6L	5H
5.7	6H	6H	6L	6L	6L	6L
5.8	6H	6H	6H	6L	6L	6L
5.9	6H	6H	6H	6H	6L	6L
6.0	6H	6H	6H	6H	6H	6L
6.1	7L	6H	6H	6H	6H	6H
6.2	7L	7L	6H	6H	6H	6H
6.3	7L	7L	7L	6H	6H	6H
6.4	7L	7L	7L	7L	6H	6H
6.5	7H	7L	7L	7L	7L	6H
6.6	7H	7H	7L	7L	7L	7L
6.7	7H	7H	7H	7L	7L	7L
6.8	7H	7H	7H	7H	7L	7L
6.9	7H	7H	7H	7H	7H	7L
7.0	8L	7H	7H	7H	7H	7H
7.1	8L	8L	7H	7H	7H	7H
7.2	8L	8L	8L	7H	7H	7H
7.3	8L	8L	8L	8L	7H	7H
7.4	8H	8L	8L	8L	8L	7H
7.5	8H	8H	8L	8L	8L	8L
7.6	8H	8H	8H	8L	8L	8L
7.7	8H	8H	8H	8H	8L	8L
7.8	8H	8H	8H	8H	8H	8L
7.9	9L	8H	8H	8H	8H	8H
8.0	9L	9L	8H	8H	8H	8H
8.1	9L	9L	9L	8H	8H	8H
8.2	9L	9L	9L	9L	8H	8H
8.3	9H	9L	9L	9L	9L	8H
8.4	9H	9H	9L	9L	9L	9L
8.5	9H	9H	9H	9L	9L	9L
8.6	9H	9H	9H	9H	9L	9L
8.7	9H	9H	9H	9H	9H	9L
8.8	10L	9H	9H	9H	9H	9H
8.9	10L	10L	9H	9H	9H	9H

TABLE 7 (Continued)

G SCORE	NUMBER OF MONTHS CLASS HAS BEEN IN HALF OF GRADE IT IS NOW IN					
G	0	1	2	3	4	5
9.0	10L	10L	10L	9H	9H	9H
9.1	10L	10L	10L	10L	9H	9H
9.2	10H	10L	10L	10L	10L	9H
9.3	10H	10H	10L	10L	10L	10L
9.4	10H	10H	10H	10L	10L	10L
9.5	10H	10H	10H	10H	10L	10L
9.6	10H	10H	10H	10H	10H	10L
9.7	11L	10H	10H	10H	10H	10H
9.8	11L	11L	10H	10H	10H	10H
9.9	11L	11L	11L	10H	10H	10H
10.0	11L	11L	11L	11L	10H	10H
10.1	11H	11L	11L	11L	11L	10H
10.2	11H	11H	11L	11L	11L	11L
10.3	11H	11H	11H	11L	11L	11L
10.4	11H	11H	11H	11H	11L	11L
10.5	11H	11H	11H	11H	11H	11L
10.6	12L	11H	11H	11H	11H	11H
10.7	12L	12L	11H	11H	11H	11H
10.8	12L	12L	12L	11H	11H	11H
10.9	12L	12L	12L	12L	11H	11H
11.0	12H	12L	12L	12L	12L	11H
11.1	12H	12H	12L	12L	12L	12L
11.2	12H	12H	12H	12L	12L	12L
11.3	12H	12H	12H	12H	12L	12L
11.4	12H	12H	12H	12H	12H	12L
11.5	13L	12H	12H	12H	12H	12H
11.6	13L	13L	12H	12H	12H	12H
11.7	13L	13L	13L	12H	12H	12H
11.8	13L	13L	13L	13L	12H	12H
11.9	13H	13L	13L	13L	13L	12H
12.0	13H	13H	13L	13L	13L	13L
12.1	13H	13H	13H	13L	13L	13L
12.2	13H	13H	13H	13H	13L	13L
12.3	13H	13H	13H	13H	13H	13L
12.4	14L	13H	13H	13H	13H	13H
12.5	14L	14L	13H	13H	13H	13H
12.6	14L	14L	14L	13H	13H	13H
12.7	14L	14L	14L	14L	13H	13H
12.8	14H	14L	14L	14L	14L	13H
12.9	14H	14H	14L	14L	14L	14L
13.0	14H	14H	14H	14L	14L	14L

TABLE 8

TABLE FOR A SCHOOL WHICH ATTEMPTS TO DO 1.0 STANDARD GRADE PER YEAR, SHOWING THE AUTOMATIC CLASSIFICATION OF PUPILS INTO GRADES ON THE BASIS OF ANY G (GRADE) SCORE

G SCORE	NUMBER OF MONTHS CLASS HAS BEEN IN HALF OF GRADE IT IS NOW IN					
G	0	1	2	3	4	5
0.0	KL	KL	KL	KL	KL	KL
0.1	KL	KL	KL	KL	KL	KL
0.2	KL	KL	KL	KL	KL	KL
0.3	KH	KL	KL	KL	KL	KL
0.4	KH	KH	KL	KL	KL	KL
0.5	KH	KH	KH	KL	KL	KL
0.6	KH	KH	KH	KH	KL	KL
0.7	KH	KH	KH	KH	KH	KL
0.8	1L	KH	KH	KH	KH	KH
0.9	1L	1L	KH	KH	KH	KH
1.0	1L	1L	1L	KH	KH	KH
1.1	1L	1L	1L	1L	KH	KH
1.2	1L	1L	1L	1L	1L	KH
1.3	1H	1L	1L	1L	1L	1L
1.4	1H	1H	1L	1L	1L	1L
1.5	1H	1H	1H	1L	1L	1L
1.6	1H	1H	1H	1H	1L	1L
1.7	1H	1H	1H	1H	1H	1L
1.8	2L	1H	1H	1H	1H	1H
1.9	2L	2L	1H	1H	1H	1H
2.0	2L	2L	2L	1H	1H	1H
2.1	2L	2L	2L	2L	1H	1H
2.2	2L	2L	2L	2L	2L	1H
2.3	2H	2L	2L	2L	2L	2L
2.4	2H	2H	2L	2L	2L	2L
2.5	2H	2H	2H	2L	2L	2L
2.6	2H	2H	2H	2H	2L	2L
2.7	2H	2H	2H	2H	2H	2L
2.8	3L	2H	2H	2H	2H	2H
2.9	3L	3L	2H	2H	2H	2H
3.0	3L	3L	3L	2H	2H	2H
3.1	3L	3L	3L	3L	2H	2H
3.2	3L	3L	3L	3L	3L	2H
3.3	3H	3L	3L	3L	3L	3L
3.4	3H	3H	3L	3L	3L	3L
3.5	3H	3H	3H	3L	3L	3L
3.6	3H	3H	3H	3H	3L	3L
3.7	3H	3H	3H	3H	3H	3L
3.8	4L	3H	3H	3H	3H	3H
3.9	4L	4L	3H	3H	3H	3H
4.0	4L	4L	4L	3H	3H	3H
4.1	4L	4L	4L	4L	3H	3H
4.2	4L	4L	4L	4L	4L	3H
4.3	4H	4L	4L	4L	4L	4L

TABLE 8 (Continued)

G SCORE	NUMBER OF MONTHS CLASS HAS BEEN IN HALF OF GRADE IT IS Now IN					
G	0	1	2	3	4	5
4.4	4H	4H	4L	4L	4L	4L
4.5	4H	4H	4H	4L	4L	4L
4.6	4H	4H	4H	4H	4L	4L
4.7	4H	4H	4H	4H	4H	4L
4.8	5L	4H	4H	4H	4H	4H
4.9	5L	5L	4H	4H	4H	4H
5.0	5L	5L	5L	4H	4H	4H
5.1	5L	5L	5L	5L	4H	4H
5.2	5L	5L	5L	5L	5L	4H
5.3	5H	5L	5L	5L	5L	5L
5.4	5H	5H	5L	5L	5L	5L
5.5	5H	5H	5H	5L	5L	5L
5.6	5H	5H	5H	5H	5L	5L
5.7	5H	5H	5H	5H	5H	5L
5.8	6L	5H	5H	5H	5H	5H
5.9	6L	6L	5H	5H	5H	5H
6.0	6L	6L	6L	5H	5H	5H
6.1	6L	6L	6L	6L	5H	5H
6.2	6L	6L	6L	6L	6L	5H
6.3	6H	6L	6L	6L	6L	6L
6.4	6H	6H	6L	6L	6L	6L
6.5	6H	6H	6H	6L	6L	6L
6.6	6H	6H	6H	6H	6L	6L
6.7	6H	6H	6H	6H	6H	6L
6.8	7L	6H	6H	6H	6H	6H
6.9	7L	7L	6H	6H	6H	6H
7.0	7L	7L	7L	6H	6H	6H
7.1	7L	7L	7L	7L	6H	6H
7.2	7L	7L	7L	7L	7L	6H
7.3	7H	7L	7L	7L	7L	7L
7.4	7H	7H	7L	7L	7L	7L
7.5	7H	7H	7H	7L	7L	7L
7.6	7H	7H	7H	7H	7L	7L
7.7	7H	7H	7H	7H	7H	7L
7.8	8L	7H	7H	7H	7H	7H
7.9	8L	8L	7H	7H	7H	7H
8.0	8L	8L	8L	7H	7H	7H
8.1	8L	8L	8L	8L	7H	7H
8.2	8L	8L	8L	8L	8L	7H
8.3	8H	8L	8L	8L	8L	8L
8.4	8H	8H	8L	8L	8L	8L
8.5	8H	8H	8H	8L	8L	8L
8.6	8H	8H	8H	8H	8L	8L
8.7	8H	8H	8H	8H	8H	8L
8.8	9L	8H	8H	8H	8H	8H
8.9	9L	9L	8H	8H	8H	8H
9.0	9L	9L	9L	8H	8H	8H

TABLE 8 (Continued)

G SCORE	NUMBER OF MONTHS CLASS HAS BEEN IN HALF OF GRADE IT IS Now In					
	G	0	1	2	3	4
9.1	9L	9L	9L	9L	8H	8H
9.2	9L	9L	9L	9L	9L	8H
9.3	9H	9L	9L	9L	9L	9L
9.4	9H	9H	9L	9L	9L	9L
9.5	9H	9H	9H	9L	9L	9L
9.6	9H	9H	9H	9H	9L	9L
9.7	9H	9H	9H	9H	9H	9L
9.8	10L	9H	9H	9H	9H	9H
9.9	10L	10L	9H	9H	9H	9H
10.0	10L	10L	10L	9H	9H	9H
10.1	10L	10L	10L	10L	9H	9H
10.2	10L	10L	10L	10L	10L	9H
10.3	10H	10L	10L	10L	10L	10L
10.4	10H	10H	10L	10L	10L	10L
10.5	10H	10H	10H	10L	10L	10L
10.6	10H	10H	10H	10H	10L	10L
10.7	10H	10H	10H	10H	10H	10L
10.8	11L	10H	10H	10H	10H	10H
10.9	11L	11L	10H	10H	10H	10H
11.0	11L	11L	11L	10H	10H	10H
11.1	11L	11L	11L	11L	10H	10H
11.2	11L	11L	11L	11L	11L	10H
11.3	11H	11L	11L	11L	11L	11L
11.4	11H	11H	11L	11L	11L	11L
11.5	11H	11H	11H	11L	11L	11L
11.6	11H	11H	11H	11H	11L	11L
11.7	11H	11H	11H	11H	11H	11L
11.8	12L	11H	11H	11H	11H	11H
11.9	12L	12L	11H	11H	11H	11H
12.0	12L	12L	12L	11H	11H	11H
12.1	12L	12L	12L	12L	11H	11H
12.2	12L	12L	12L	12L	12L	11H
12.3	12H	12L	12L	12L	12L	12L
12.4	12H	12H	12L	12L	12L	12L
12.5	12H	12H	12H	12L	12L	12L
12.6	12H	12H	12H	12H	12L	12L
12.7	12H	12H	12H	12H	12H	12L
12.8	13L	12H	12H	12H	12H	12H
12.9	13L	13L	12H	12H	12H	12H
13.0	13L	13L	13L	12H	12H	12H

TABLE 9

TABLE FOR A SCHOOL WHICH ATTEMPTS TO DO 1.1 STANDARD GRADES PER YEAR, SHOWING THE AUTOMATIC CLASSIFICATION OF PUPILS INTO GRADES ON THE BASIS OF ANY G (GRADE) SCORE ¹

G SCORE	NUMBER OF MONTHS CLASS HAS BEEN IN HALF OF GRADE IT IS NOW IN						
	G	0	1	2	3	4	5
0.0	KL	KL	KL	KL	KL	KL	KL
0.1	KL	KL	KL	KL	KL	KL	KL
0.2	KL	KL	KL	KL	KL	KL	KL
0.3	KH	KL	KL	KL	KL	KL	KL
0.4	KH	KH	KL	KL	KL	KL	KL
0.5	KH	KH	KH	KL	KL	KL	KL
0.6	KH	KH	KH	KH	KL	KL	KL
0.7	KH	KH	KH	KH	KH	KH	KL
0.8	KH	KH	KH	KH	KH	KH	KH
0.9	1L	KH	KH	KH	KH	KH	KH
1.0	1L	1L	KH	KH	KH	KH	KH
1.1	1L	1L	1L	KH	KH	KH	KH
1.2	1L	1L	1L	1L	KH	KH	KH
1.3	1L	1L	1L	1L	1L	1L	KH
1.4	1L	1L	1L	1L	1L	1L	1L
1.5	1H	1L	1L	1L	1L	1L	1L
1.6	1H	1H	1L	1L	1L	1L	1L
1.7	1H	1H	1H	1L	1L	1L	1L
1.8	1H	1H	1H	1H	1L	1L	1L
1.9	1H	1H	1H	1H	1H	1H	1L
2.0	2L	1H	1H	1H	1H	1H	1H
2.1	2L	2L	1H	1H	1H	1H	1H
2.2	2L	2L	2L	1H	1H	1H	1H
2.3	2L	2L	2L	2L	1H	1H	1H
2.4	2L	2L	2L	2L	2L	2L	1H
2.5	2L	2L	2L	2L	2L	2L	2L
2.6	2H	2L	2L	2L	2L	2L	2L
2.7	2H	2H	2L	2L	2L	2L	2L
2.8	2H	2H	2H	2L	2L	2L	2L
2.9	2H	2H	2H	2H	2L	2L	2L
3.0	2H	2H	2H	2H	2H	2H	2L
3.1	3L	2H	2H	2H	2H	2H	2H
3.2	3L	3L	2H	2H	2H	2H	2H
3.3	3L	3L	3L	2H	2H	2H	2H
3.4	3L	3L	3L	3L	2H	2H	2H
3.5	3L	3L	3L	3L	3L	3L	2H
3.6	3L	3L	3L	3L	3L	3L	3L
3.7	3H	3L	3L	3L	3L	3L	3L
3.8	3H	3H	3L	3L	3L	3L	3L
3.9	3H	3H	3H	3L	3L	3L	3L
4.0	3H	3H	3H	3H	3L	3L	3L
4.1	3H	3H	3H	3H	3H	3H	3L
4.2	4L	3H	3H	3H	3H	3H	3H

¹ Table prepared with the assistance of Grace Moffatt.

TABLE 9 (Continued)

G SCORE	NUMBER OF MONTHS CLASS HAS BEEN IN HALF OF GRADE IT IS Now In					
	G	0	1	2	3	4
4.3	4L	4L	3H	3H	3H	3H
4.4	4L	4L	4L	3H	3H	3H
4.5	4L	4L	4L	4L	3H	3H
4.6	4L	4L	4L	4L	4L	3H
4.7	4L	4L	4L	4L	4L	4L
4.8	4H	4L	4L	4L	4L	4L
4.9	4H	4H	4L	4L	4L	4L
5.0	4H	4H	4H	4L	4L	4L
5.1	4H	4H	4H	4H	4L	4L
5.2	4H	4H	4H	4H	4H	4L
5.3	5L	4H	4H	4H	4H	4H
5.4	5L	5L	4H	4H	4H	4H
5.5	5L	5L	5L	4H	4H	4H
5.6	5L	5L	5L	5L	4H	4H
5.7	5L	5L	5L	5L	5L	4H
5.8	5L	5L	5L	5L	5L	5L
5.9	5H	5L	5L	5L	5L	5L
6.0	5H	5H	5L	5L	5L	5L
6.1	5H	5H	5H	5L	5L	5L
6.2	5H	5H	5H	5H	5L	5L
6.3	5H	5H	5H	5H	5H	5L
6.4	6L	5H	5H	5H	5H	5H
6.5	6L	6L	5H	5H	5H	5H
6.6	6L	6L	6L	5H	5H	5H
6.7	6L	6L	6L	6L	5H	5H
6.8	6L	6L	6L	6L	6L	5H
6.9	6L	6L	6L	6L	6L	6L
7.0	6H	6L	6L	6L	6L	6L
7.1	6H	6H	6L	6L	6L	6L
7.2	6H	6H	6H	6L	6L	6L
7.3	6H	6H	6H	6H	6L	6L
7.4	6H	6H	6H	6H	6H	6L
7.5	7L	6H	6H	6H	6H	6H
7.6	7L	7L	6H	6H	6H	6H
7.7	7L	7L	7L	6H	6H	6H
7.8	7L	7L	7L	7L	6H	6H
7.9	7L	7L	7L	7L	7L	6H
8.0	7L	7L	7L	7L	7L	7L
8.1	7H	7L	7L	7L	7L	7L
8.2	7H	7H	7L	7L	7L	7L
8.3	7H	7H	7H	7L	7L	7L
8.4	7H	7H	7H	7H	7L	7L
8.5	7H	7H	7H	7H	7H	7L
8.6	8L	7H	7H	7H	7H	7H
8.7	8L	8L	7H	7H	7H	7H
8.8	8L	8L	8L	7H	7H	7H
8.9	8L	8L	8L	8L	7H	7H

HOW TO CLASSIFY PUPILS

TABLE 9 (Continued)

G SCORE	NUMBER OF MONTHS CLASS HAS BEEN IN HALF OF Now In				
G	0	1	2	3	4
9.0	8L	8L	8L	8L	8L
9.1	8L	8L	8L	8L	8L
9.2	8H	8L	8L	8L	8L
9.3	8H	8H	8L	8L	8L
9.4	8H	8H	8H	8L	8L
9.5	8H	8H	8H	8H	8L
9.6	8H	8H	8H	8H	8H
9.7	9L	8H	8H	8H	8H
9.8	9L	9L	8H	8H	8H
9.9	9L	9L	9L	8H	8H
10.0	9L	9L	9L	9L	8H
10.1	9L	9L	9L	9L	9L
10.2	9L	9L	9L	9L	9L
10.3	9H	9L	9L	9L	9L
10.4	9H	9H	9L	9L	9L
10.5	9H	9H	9H	9L	9L
10.6	9H	9H	9H	9H	9L
10.7	9H	9H	9H	9H	9H
10.8	10L	9H	9H	9H	9H
10.9	10L	10L	9H	9H	9H
11.0	10L	10L	10L	9H	9H
11.1	10L	10L	10L	10L	9H
11.2	10L	10L	10L	10L	10L
11.3	10L	10L	10L	10L	10L
11.4	10H	10L	10L	10L	10L
11.5	10H	10H	10L	10L	10L
11.6	10H	10H	10H	10L	10L
11.7	10H	10H	10H	10H	10L
11.8	10H	10H	10H	10H	10H
11.9	11L	10H	10H	10H	10H
12.0	11L	11L	10H	10H	10H
12.1	11L	11L	11L	10H	10H
12.2	11L	11L	11L	11L	10H
12.3	11L	11L	11L	11L	11L
12.4	11L	11L	11L	11L	11L
12.5	11H	11L	11L	11L	11L
12.6	11H	11H	11L	11L	11L
12.7	11H	11H	11H	11L	11L
12.8	11H	11H	11H	11H	11L
12.9	11H	11H	11H	11H	11H
13.0	12L	11H	11H	11H	11H

column 0. A period of more than fifteen calendar days is counted as a month; for example, if a class has been in a grade for three months and sixteen days, readings will be made in column 4.

To illustrate, Pupil 1 of Table 4 has a Gp of 2.9. We find 2.9 in the G column of Table 9 and read the symbol opposite 2.9 in the 0 column. We find 2H. We therefore write 2H in the statistical classification column of Table 4. Pupil 4 has a Gp of 4.1. We find 4.1 in the G column. Opposite 4.1 in the 0 column, we find 3H. We therefore write 3H in the statistical classification column of Table 4.

Let us suppose that our tests were administered on March 10. Even though the school system is on the annual promotion basis, Classification Table 9 is usable. In this case the classes would have been in the second half of the grade a little more than one month. We would therefore read in the column headed 1. Suppose a pupil has a Gp of 4.3. Then the statistical classification of this pupil is 4L.

It must not be supposed that one may read directly from the Gp the grade into which a pupil should go and thus avoid the use of the classification table. To illustrate, if the school is one that attempts to accomplish 1.1 standard grades work per year, it might be assumed that a Gp of 4.2 would place a pupil in the fourth grade. This is not necessarily true. For example, if tests are given during the first fifteen calendar days of school, according to Table 9, the reading is 4L. But if the tests are given during any other part of the semester, the reading is 3H, and the pupil would therefore be classified in the third grade. The same situation exists in a school which attempts to do one standard grade per year. From Table 8 it is seen that if a Gp is 4.2 and the class has been in the half of the grade 0, 1, 2, 3, or 4 months, the reading is 4L. But if the class has been in the grade 5 months, the reading is 3H.

Step 14. Determine Conservative Classification.—It may be neither feasible nor desirable to adhere strictly to the statistical classification in the placement of pupils. Among the situations which may arise are the following:

The grade in which a pupil is placed by the statistical classification may not exist in the school organization. In Table 4, Pupil 2 has a statistical classification of 3H. The school is on the annual promotion basis. The only third grade in the school

is 3L, that is, the pupils are just beginning third-grade work. The question then arises: Shall Pupil 2 be placed in the third grade or in the fourth grade? A technique for answering this and similar questions is therefore needed.

A similar situation occurs in school systems which have six grades in the elementary school, followed by a junior high school which includes Grades VII, VIII, and IX. A pupil, for example, who is in the 6L grade may have a statistical classification of 8L. It may be unwise for this pupil to skip the work of both the sixth and the seventh grades. Indeed, since the junior high school is under another principal, it is obviously impossible for the elementary principal to promote the pupil to the 8L grade. Similar cases will be found in every elementary school, whether it has six, seven, eight, or nine grades.

Again, the statistical classification may cause many radical changes, particularly if standard tests have not previously been used in classification. It is better to be conservative at first and thus retain the confidence and coöperation of teachers and parents.

If, therefore, one wishes to be conservative, the technique to be followed in obtaining a conservative classification is as follows:

a. Determine the classification standards for the grades immediately above and below the group to be classified. The standards must be appropriate to the classification table that is used. If the 0.9 Classification Table is used, the appropriate standard will be found in Table 10. For example, if the grade is 3L, and if the tests are given on September 15, we find the classification standards for 2L and 4L to be 1.9 and 3.7 respectively. In a school promoting semi-annually, the classification standards would be 2.4 and 3.3.

If the 1.1 Classification Table is used, the appropriate classification standards will be found in Table 12. For example, if the tests are given on March 20, in a yearly system, the third grade is then 3H, and the desired classification standards are 2.9 and 5.1 in an annual system.

b. Keeping these classification standards in mind, we return to the class whose pupils are to be classified, and find all the Gp scores which are larger than the classification standard of the grade above. In the Conservative Classification column, oppo-

site these scores should be written the symbol for this next higher grade.

To illustrate, in Table 4 the grade to be classified is the 3L. The grade just above is 4L, since this is a school which promotes annually. The classification standard for Grade 4L may be found in Table 12. Since the date of testing is September 9, we will read in the first column. Opposite 4L we find 4.3. Pupils 7, 8, 12, and 14 have Gp scores larger than 4.3. Opposite their names, therefore, we find 4L in the Conservative Classification column.

TABLE 10
CLASSIFICATION STANDARDS
FOR A SCHOOL USING THE 0.9 CLASSIFICATION TABLE

GRADE	AUG. 16 TO SEPT. 15	SEPT. 16 TO OCT. 15	OCT. 16 TO NOV. 15	NOV. 16 TO DEC. 15	DEC. 16 TO JAN. 15
	JAN. 16 TO FEB. 15	FEB. 16 TO MAR. 15	MAR. 16 TO APR. 15	APR. 16 TO MAY 15	MAY 16 TO JUNE 15
1L.....	1.0	1.1	1.2	1.3	1.4
1H.....	1.5	1.6	1.7	1.8	1.9
2L.....	1.9	2.0	2.1	2.2	2.3
2H.....	2.4	2.5	2.6	2.7	2.8
3L.....	2.8	2.9	3.0	3.1	3.2
3H.....	3.3	3.4	3.5	3.6	3.7
4L.....	3.7	3.8	3.9	4.0	4.1
4H.....	4.2	4.3	4.4	4.5	4.6
5L.....	4.6	4.7	4.8	4.9	5.0
5H.....	5.1	5.2	5.3	5.4	5.5
6L.....	5.5	5.6	5.7	5.8	5.9
6H.....	6.0	6.1	6.2	6.3	6.4
7L.....	6.4	6.5	6.6	6.7	6.8
7H.....	6.9	7.0	7.1	7.2	7.3
8L.....	7.3	7.4	7.5	7.6	7.7
8H.....	7.8	7.9	8.0	8.1	8.2
9L.....	8.2	8.3	8.4	8.5	8.6
9H.....	8.7	8.8	8.9	9.0	9.1
10L.....	9.1	9.2	9.3	9.4	9.5
10H.....	9.6	9.7	9.8	9.9	10.0
11L.....	10.0	10.1	10.2	10.3	10.4
11H.....	10.5	10.6	10.7	10.8	10.9
12L.....	10.9	11.0	11.1	11.2	11.3
12H.....	11.4	11.5	11.6	11.7	11.8

In effect, this procedure insures that a pupil is given a special promotion only if his Gp exceeds the classification standard, or the approximate mean, of the grade to which he goes. Furthermore, all special promotions are promotions to the grade just

TABLE 11
CLASSIFICATION STANDARDS
FOR A SCHOOL USING THE 1.0 CLASSIFICATION TABLE

GRADE	AUG 16 TO SEPT 15	SEPT. 16 TO OCT 15	OCT. 16 TO NOV. 15	NOV. 16 TO DEC. 15	DEC. 16 TO JAN. 15
	JAN 16 TO FEB. 15	FEB 16 TO MAR 15	MAR 16 TO APR 15	APR. 16 TO MAY 15	MAY 16 TO JUNE 15
1L	1.0	1.1	1.2	1.3	1.4
1H	1.5	1.6	1.7	1.8	1.9
2L	2.0	2.1	2.2	2.3	2.4
2H	2.5	2.6	2.7	2.8	2.9
3L	3.0	3.1	3.2	3.3	3.4
3H	3.5	3.6	3.7	3.8	3.9
4L	4.0	4.1	4.2	4.3	4.4
4H	4.5	4.6	4.7	4.8	4.9
5L	5.0	5.1	5.2	5.3	5.4
5H	5.5	5.6	5.7	5.8	5.9
6L	6.0	6.1	6.2	6.3	6.4
6H	6.5	6.6	6.7	6.8	6.9
7L	7.0	7.1	7.2	7.3	7.4
7H	7.5	7.6	7.7	7.8	7.9
8L	8.0	8.1	8.2	8.3	8.4
8H	8.5	8.6	8.7	8.8	8.9
9L	9.0	9.1	9.2	9.3	9.4
9H	9.5	9.6	9.7	9.8	9.9
10L	10.0	10.1	10.2	10.3	10.4
10H	10.5	10.6	10.7	10.8	10.9
11L	11.0	11.1	11.2	11.3	11.4
11H	11.5	11.6	11.7	11.8	11.9
12L	12.0	12.1	12.2	12.3	12.4
12H	12.5	12.6	12.7	12.8	12.9

above the grade in which a pupil is seated. This means that a pupil, even in a yearly system, will "skip," or miss, at most only one year's instruction.

c. To determine the conservative classification for pupils with a low Gp, find in the Gp column all the scores which are smaller than the classification standard of the grade just below. In the Conservative Classification column, opposite these scores, the symbol for the grade just below should be written.

In Table 4 the grade under consideration is 3L and the classification standard for the grade just below, which is 2L, is 2.1. No pupils have scores below 2.1. If there were any, we would write 2L in the Conservative Classification column opposite their names.

TABLE 12
CLASSIFICATION STANDARDS
FOR A SCHOOL USING THE 1.1 CLASSIFICATION TABLE

GRADE	AUG 16 TO SEPT. 15	SEPT 16 TO OCT 15	OCT 16 TO NOV. 15	NOV 16 TO DEC 15	DEC 16 TO JAN 15
	JAN. 16 TO FEB 15	FEB. 16 TO MAR 15	MAR 16 TO APR 15	APR 16 TO MAY 15	MAY 16 TO JUNE 15
1L	1.0	1.1	1.2	1.3	1.4
1H	1.6	1.7	1.8	1.9	2.0
2L	2.1	2.2	2.3	2.4	2.5
2H	2.7	2.8	2.9	3.0	3.1
3L	3.2	3.3	3.4	3.5	3.6
3H	3.8	3.9	4.0	4.1	4.2
4L	4.3	4.4	4.5	4.6	4.7
4H	4.9	5.0	5.1	5.2	5.3
5L	5.4	5.5	5.6	5.7	5.8
5H	6.0	6.1	6.2	6.3	6.4
6L	6.5	6.6	6.7	6.8	6.9
6H	7.1	7.2	7.3	7.4	7.5
7L	7.6	7.7	7.8	7.9	8.0
7H	8.2	8.3	8.4	8.5	8.6
8L	8.7	8.8	8.9	9.0	9.1
8H	9.3	9.4	9.5	9.6	9.7
9L	9.8	9.9	10.0	10.1	10.2
9H	10.4	10.5	10.6	10.7	10.8
10L	10.9	11.0	11.1	11.2	11.3
10H	11.5	11.6	11.7	11.8	11.9
11L	12.0	12.1	12.2	12.3	12.4
11H	12.6	12.7	12.8	12.9	13.0
12L	13.1	13.2	13.3	13.4	13.5
12H	13.7	13.8	13.9	14.0	14.1

In effect, this procedure does not demote a pupil more than one grade, and then only if his Gp is below the classification standard, or the approximate mean, of the grade below.

d. For pupils whose Gp is smaller than the classification standard of the grade above and larger than the classification standard of the grade below, the conservative classification coincides with their present grade placement. On the Class Record Sheet, therefore, the symbol for the present grade should be written in the Conservative Classification column opposite their names. In Table 4 we find 3L opposite the names of these pupils. At this point the question may be raised: Why determine statistical classification if we are going to use the conservative classification? The statistical classification shows clearly where the pupil *ought* to be, though we may wish to be much more

conservative in our actual placement. For example, in Table 4 there are four pupils, Nos. 7, 8, 12, and 14, who have a statistical classification of 4H. Although a conservative policy would not place these pupils higher than 4L, it is well to keep a record of their possibilities and where they belong by attainment.

Step 15. Determine Actual Classification.—We are now ready to take final action and determine *actual* classification. It will be observed that the Class Record Sheet in Table 4 does not contain such a column. In some schools, the principal and teachers may wish to follow the conservative classification. After a school has been using tests as a basis of classification, however, it may be and often is desirable to follow the statistical classification more closely.

Suppose that a definite policy in this matter has been agreed upon. There is still another step to be taken. The proposed classification of each pupil should be scrutinized in the light of all factors which might have a bearing on the problem of classification. Among these factors are chronological age, physiological maturity, social maturity, brightness, dependability, health, judgment of parents, and the like. It is, of course, impossible to determine in advance what weight each of these factors should have.

Again, it may be desirable to examine a pupil's subject profile, that is, his achievement in each separate subject, as compared with his Gp. A graph will reveal unevenness of achievement. For example, a pupil's Gp may be just sufficient to warrant an extra promotion. He may, however, be deficient in reading, as indicated by his Gr. Since this is a highly important skill—since his mastery of other subjects depends largely upon his ability in reading—this pupil might be seriously handicapped.

In general, there will be only a few cases in which it will be desirable to make the actual classification differ from the conservative classification previously obtained. In the six grades of our sample school, for example, the conservative classification was followed in all except six cases. These six cases are described in Table 13. The complete Class Record Sheet for the 3L grade of the sample school is found in Table 4. The Class Record Sheets for the other grades are not printed.

A special situation in which the actual classification differs from that obtained by previous computations occurs when a

TABLE 13
SPECIAL CASES IN THE SAMPLE SCHOOL IN WHICH THE CONSERVATIVE CLASSIFICATION WAS NOT FOLLOWED

GRADE	PUPIL No	Gp	CON-SERV. CLASS.	ACTUAL CLASS	REASON
3L	3	4.2	3L	4L	Pupil has high Gi and Gr. The Gp is just 0.2 below the score required for a conservative classification of 4L.
4L	42	3.2	4L	3L	Gp is just 0.1 above the score required for a conservative classification of 3L. His Gi (2.5) is low, and he is comparatively young and immature.
4L	40	5.5	5L	4L	Pupil's chronological age is 8-7, which is considerably less than that of any pupil in 5L. Although he is physiologically as mature as the average eight-year-old, he might be out of place socially in the older group.
5L	47	6.3	5L	6L	Pupil's Gp is almost as high as the classification standard of 6L. His chronological age (12-11) is far above the mean of 5L; in fact, only two pupils in 6L are older than he is. He is correspondingly mature socially.
5L	66	6.7	6L	5L	Pupil is young (9-6). He is not very evenly developed, as indicated by variation in G scores (Gi 6.8; Gr 6.7; Ga 5.6; Gs 7.1). Finally, he is in frail health.
6L	86	7.5	6L	7L	Pupil is fairly old (12-5). His low Gr (4.9) is due to a language handicap. His mother is very anxious for him to get the extra promotion, as it will be necessary for him to go to work to help support the family as soon as the law permits.

school wishes to wait until the end of the semester or year to make promotions and demotions. This is particularly true when tests are given near the end of the semester or year. The recommended procedure is as follows: Determine, according to the principles just laid down, the actual classification as of the

date of testing. These figures show the pupil classification as it should be for *that* year (or semester). For the next year or semester, obviously every pupil should be promoted to the next higher grade. Therefore, the actual classification for all pupils should be changed by raising the figures one grade. For example, 4H will be changed to 5L, 3L to 3H. At the end of the semester or year, promotions will be made accordingly. The situation is exactly the same as it would be if the reclassification had taken place earlier. No actual demotions take place since the pupils who earlier in the year would have been demoted simply remain in and repeat the grade in which they are.

Does the Typical School without a System of Grouping Its Pupils Need Reclassifying?—Teachers and tests are in substantial agreement that pupils are not classified in homogeneous

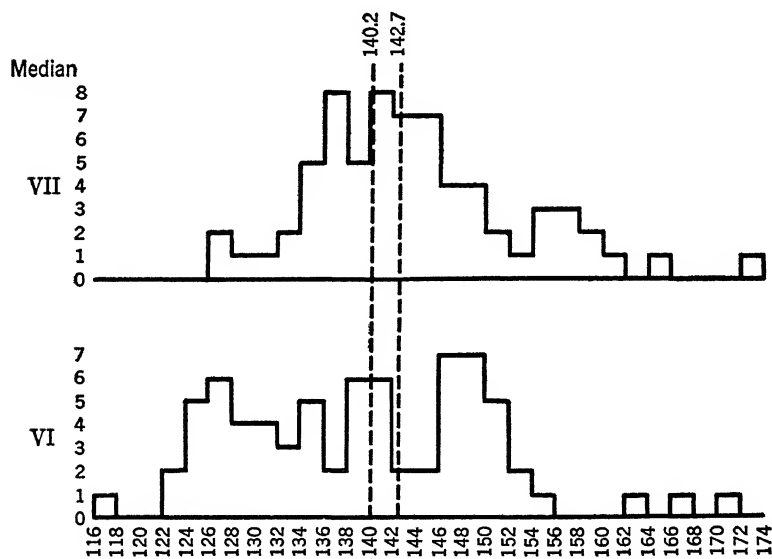


Fig. 1. A Graphic Picture of the Amount of Overlapping of the Educational Ages of Pupils in Grades VI and VII of School Y.

groups. The median educational age for each grade and section in a certain School Y was as follows:

IIIL	IIIH	IVL	IVH	VL	VH	VIL	VIH	VIIIL	VIIH	VIIIL	VIIH
100	107	112	122	128	133	143	132	144	144	149	157

In School Y, VIH is actually behind VIL and there is practically no progress at all between VIL and VIIH.

But the position of the grade medians, improperly spaced as they are, does not begin to suggest how bad the classification really is. Figure 1 permits a comparison of the amount of total grade overlapping. At a glance this diagram tells us that the extreme range of each grade is about 50 months in terms of educational age, which is equivalent to a range of about four typical grades, while the interval between the two grades is 2.5 months. The range of ability within one grade of School Y is then approximately 20 times the difference between two adjoining grades.

The Amount of Promotion and Demotion Necessary.—The amount of reclassification necessary in another School X, even when the classification has been somewhat conservative, is shown in Table 14.

TABLE 14
DISTRIBUTION OF CHANGES MADE IN RECLASSIFYING SCHOOL X BY MEANS
OF EDUCATIONAL TESTS

AMOUNT OF CHANGE	NUMBER OF PUPILS						Total
	III	IV	V	VI	VII	VIII	
Demoted Three Grades..	0	0	0	0	0	0	0
Demoted Two Grades....	0	0	0	2	2	0	5
Demoted One Grade. . .	2	1	1	1	3	3	11
No Change.	13	13	11	8	7	5	57
Promoted One Grade . .	4	1	2	5	5	1	18
Promoted Two Grades..	0	2	0	1	1	2	6
Promoted Three Grades	0	0	0	1	1	0	2

Table 15 gives similar data for a larger school—School Y—where the technique of reclassification was practically the same. The tests used in this school were reading, vocabulary, spelling, language, reasoning arithmetic, mixed fundamentals, and composition.

Schools X and Y have not been chosen because they illustrate dramatically the need for reclassification. On the contrary, they are quite typical. The conventional methods of classifying pupils are so crude that no one should regard them as sacrosanct or be overcritical of those who seek better methods.

TABLE 15

DISTRIBUTION OF CHANGES MADE IN RECLASSIFYING SCHOOL Y BY MEANS OF EDUCATIONAL TESTS

AMOUNT OF CHANGE	NUMBER OF PUPILS						Total
	III	IV	V	VI	VII	VIII	
Demoted Four Grades...	0	0	0	0	0	0	0
Demoted Three Grades..	0	0	0	0	0	0	0
Demoted Two Grades...	0	0	0	0	0	1	1
Demoted One Grade....	0	4	1	2	5	13	25
No Change.....	39	39	38	29	45	36	226
Promoted One Grade....	15	16	39	27	12	9	118
Promoted Two Grades...	0	5	9	11	3	1	29
Promoted Three Grades..	0	0	0	1	1	3	5
Promoted Four Grades..	0	0	0	2	0	1	3

The two tables when combined lead to the following conclusions which need to be only slightly discounted because of unreliability of the tests:

1. About 44 per cent of pupils are wrongly classified.
2. About 34 per cent of pupils are misplaced one grade.
3. About 10 per cent of pupils are misplaced two or more grades.
4. Only about 8 per cent of pupils are pushed ahead of the grade where they belong, while nearly 36 per cent are held back from the grade where they belong.

The reclassification of the pupils in School X was recommended on the following conditions: (1) All promotions and demotions were to be trial promotions and demotions, and the pupils were to be so informed. (2) After four weeks of trial, the principal in consultation with the teachers was to construct a series of examinations upon the material studied during the four weeks and try these tests upon the pupils. (3) The teachers were to rank the pupils in their respective classes upon the quality of their work during the four weeks. (4) The principal and teachers were to decide the final disposition of the pupils.

The demoted pupils have "made good," i.e., there has been no disposition to question the recommendations in their cases. Not one has been returned to his original grade. What happened to the promoted pupils for whom reports are available is shown in Table 16. This table is read as follows: Pupil

War J., who has an E.Q. of 90, was promoted over Grade IV. He ranked, according to the educational tests, first among the sixteen pupils who, together with him, made up Grade V. He ranked first among the same sixteen pupils who were tested

TABLE 16

WHAT HAPPENED TO THE SPECIALLY PROMOTED PUPILS OF SCHOOL X

PUPIL	E Q	GRADE SKIPPED	RANK BY ED AGE	RANK BY PRIN- CIPAL	RANK BY TEACHER	FINAL DISPO- SITION
War J. . . .	90	IV	1-16	1-16	6-16	V
War R. . . .	108	IV	4-16	2-16	3-16	V
Kim. . . .	134	IV	10-16	7-16	9-16	IV
Ant	112	IV	12-16	8-16	11-16	IV
Mye.	114	V & VI	1-16	3-16	5-16	VII
Sco.	125	VI	3-16	11-16	14-16	VI
Van.	110	V & VI	7-16	—	13-16	VI
Hoy.	112	VI	10-16	7-16	9-16	VII
Fra.	135	VII	1-16	4-16	6-16	VIII
Kim	131	VII	4-16	10-16	9-16	VIII
Spi.	118	VII	7-16	14-16	13-16	VII
Lan	108	VII	9-16	16-16	16-16	VII
Pug.	110	VII	10-16	12-16	14-16	VII
Ant.	121	VII	11-16	—	11-16	VII
Mit.	113	VII	12-16	8-16	15-16	VII
Pug.	126	VIII	3-10	1-10	5-10	IX
Average.			6.2-15.6	7.4-15.6	9.6-15.6	

by the principal upon four weeks of school work. In the judgment of his teacher he ranked sixth. He was finally retained in Grade V.

Table 16 permits an interesting psychological study of the pedagogical mind. The table suggests the following:

a. A specially promoted pupil tends to be ranked lower by the teacher's judgment than by the principal's examination or by standard educational tests. The averages at the bottom of the table show that the average ranks by tests, principal, and teacher are respectively 6.2, 7.4, and 9.6 out of about sixteen pupils.

b. A young, specially promoted pupil must succeed beyond a shadow of doubt or he will be demoted. Pupils Kim and Ant of Grade V, and possibly Mit of Grade VIII, did

better than was originally anticipated and yet they were reduced a grade.

c. A pupil's educational age and E.Q. must at least exceed the median of the grade to which he is sent or the teacher and principal will probably return him. And it should be remembered that the principal and teachers of School X were friendly to the experiment.

d. The school's staff is convicted of injustice by its own measurements. Can anyone unacquainted with school traditions give a rational explanation of why pupils Kim and Ant were sent back to Grade IV? The real fact is that these teachers require a young pupil to do, not the typical work of the grade, but the *best work* in the grade. The teachers of School X testify that most of the young pupils demoted had rapidly risen in rank since the opening of school. With their high E.Q.'s it is probable that this process would continue throughout the year, thus making their class status better and better.

The teachers explained that pupils Kim and Ant were demoted, even when their rank was satisfactory, because "those ranking below them were relatively stupid pupils." This factor would not have influenced the teachers had anyone been present to explain that while these pupils were dull, they were also much over age. Additional years of schooling had balanced their dullness. While their E.Q.'s were low their educational ages were as high as pupils Kim and Ant. If the measurements of the principal and the judgments of the teachers be accepted at their face value, only one pupil was legitimately sent back and that is pupil Lan. All the others have paid the penalty of their prominence and particularly of their unfortunate youthfulness, unless it be assumed that the fundamental basis for the classification of pupils should be chronological age.

What happened to the pupils who were sent back? If the effect of demotion is to produce sulkers, special promotion should not be given unless there is considerable certainty that the promotion will be maintained. The principal reports that one or two were glad to go back to their former companions, some did not want to go back, some didn't care, every pupil except Lan are at the top or near the top of the grades to

which they were returned. The principal reports that those originally demoted on the basis of the tests are happy and satisfied.

Frazen tried the experiment in the Garden City elementary school of giving special promotion only to those pupils whose educational age and E.Q. or mental age and I.Q. both exceed the median of the grade to which they are sent. In no case was a pupil afterward demoted.

Step 16. Diagnose Strengths and Deficiencies.—Pupils are grouped that they may be taught, and diagnosis should precede teaching. But, since the discussion of the problems and procedures involved in classifying a school has already made Book Three overlong, the diagnostic interpretation of the test records in Tables 4 and 17 will be reserved for Chapters XV and XIX.

CHAPTER XII

HOW TO CLASSIFY WHEN PROMOTIONS ARE SEMI-ANNUAL

Table 17 illustrates classification in a semi-annual system. Several new features are introduced. The 1L and 1H grades are given as illustrations. The tests used in these grades are different from those used in the upper grades, and the Class Record Sheets differ accordingly. A portion of the Class Record Sheet of 3L is also given; sheets for the other grades are similar in form. A new G score, Gt, or G according to teacher's rank, will also be found in this table. The Gt may be used in a yearly system, too, but to simplify the presentation it was omitted in Table 4.

The entire procedure for classification will be summarized in the succeeding paragraphs.

Only the new features will be described.

Step 1. Compute and Record Gt.—Gt, or grade score according to teacher's rank, is frequently a useful score. Teacher's rank is, of course, comparatively valueless at the beginning of the school year.

A school in which standard tests are a part of the regular program may definitely plan, as a matter of routine, to administer a battery of tests at the end of the semester or the year. Except for newcomers (who may be tested on entrance), the school will then be satisfactorily classified when school opens. This permits the determination of Gt when it is most significant; that is, when it is based on observations of at least one semester.

The procedure in computing Gt is as follows:

a. Before she inspects the test results, the teacher should list in order of her estimate of their fitness for promotion the names of the pupils in her class. It is desirable that this ranking be made before the test results are known, since the chief value of a teacher's estimate lies in the fact that it includes qualities other than those measured by intelligence and achievement tests. Arrange the names, as listed, on a work sheet, or a piece of ruled paper. In order that the estimates of the teachers in the

TABLE 17
CLASS RECORD SHEETS
FOR SCHOOL HAVING SEMI-ANNUAL PROMOTION
A. GRADE 1L

*Franklin School**November 9**Miss A.**Room 116*

PUPIL No.	G GRADE	G AGE	Gi	Gt	Gp	CLASSIFICATION STAT. CONSERV.	
1	1.2	0.2	1.5	1.4	1.5	1H	1L
2	1.2	0.6	0.6	0.5	0.6	KH	KH
3	1.2	0.6	1.4	0.9	1.2	1L	1L
4	1.2	0.5	0.9	0.6	0.8	KH	1L
5	1.2	0.7	1.5	1.3	1.4	1L	1L
6	1.2	1.2	1.3	1.5	1.4	1L	1L
7	1.2	0.7	2.2	1.9	2.1	2L	1H
8	1.2	0.4	1.5	1.5	1.5	1H	1L
9	1.2	1.1	1.0	1.0	1.0	1L	1L
10	1.2	0.3	0.6	0.6	0.6	KH	KH
11	1.2	0.7	0.7	0.7	0.7	KH	1L
12	1.2	0.7	0.9	0.9	0.9	KH	1L
13	1.2	0.5	0.5	0.1	0.3	KL	KH
14	1.2	0.6	0.4	0.8	0.6	KH	KH
15	1.2	0.9	0.9	0.8	0.9	KH	1L
16	1.2	1.0	1.0	0.9	1.0	1L	1L
17	1.2	0.5	0.8	0.9	0.9	KH	1L
18	1.2	0.8	0.8	1.5	1.2	1L	1L
19	1.2	0.9	0.9	1.0	1.0	1L	1L
20	1.2	1.4	1.9	2.4	2.2	2L	1H
21	1.2	1.3	2.4	2.2	2.3	2L	1H
22	1.2	0.5	0.1	0.2	0.2	KL	KH
23	1.2	0.6	0.2	0.3	0.3	KL	KH
24	1.2	0.5	0.5	0.4	0.5	KH	KH
25	1.2	0.7	0.3	0.5	0.4	KL	KH
Total	30.0	17.9	24.8	24.8	25.5		
Mean	1.2	0.7	1.0		1.0		

B. GRADE 1H

*Franklin School**November 9**Miss B.**Room 118*

PUPIL No.	G GRADE	G AGE	Gi	Gt	Gr1	Gr2	Ge	Gp	CLASSIFICATION STAT. CONSERV.	
26	1.7	1.3	1.8	1.4	0.9	1.3	1.1	1.4	1L	1H
27	1.7	0.8	1.4	0.9	1.0	1.0	1.0	1.1	1L	1L
28	1.7	1.5	1.9	2.5	2.4	2.5	2.5	2.3	2L	2L
..
Total	35.7	27.3	40.0		29.0	32.4	32.2	38.5		
Mean	1.7	1.3	1.7		1.5	1.5	1.5	1.6		

TABLE 17 (Continued)

C. GRADE 3L

Franklin School

November 9

Miss E.

Room 103

PUPIL No.	G GRADE	G AGE	Gi	Gt	Gr	Ga	Ge	Gp	CLASSIFICATION	
									STAT	CONSERV.
102	3.2	2.8	2.8	2.8	.	.	.	2.8	2H	3L
103	3.2	2.5	3.7	3.9	2.8	4.3	3.6	3.7	3H	3L
104	3.2	3.2	3.7	3.6	3.7	2.2	3.0	3.4	3L	3L
..
Total	64.0	50.0	99.1		78.8	92.8	85.7	89.5		
Mean	3.2	2.5	3.6		3.2	3.7	3.4	3.6		

same school or school system may be based on like factors, it may be desirable for a group of teachers to meet and determine what qualities shall be taken into consideration. They will surely wish to be guided in part by marks which they have given each pupil on examinations, recitations, assignments, and the like, unless marks are already in G score form as advised in Book Six and are separately available for inclusion in Gp.

b. No two pupils should be given the same position. A decision should be forced in some way; if need be, by tossing a coin.

c. It should be decided whether Gi or Ge is the better measure of fitness to do the work of the next grade. This will depend on the number and character of the tests used. In our sample school (Table 17) only two achievement tests were administered. After considering the relative value of Gi and Ge, it was agreed that Gi would be the better measure of fitness for promotion. In the case of the sample school in Table 4, however, Ge might be selected as the better measure, since the number of achievement tests which are combined in Ge in that table make it a more valid measure of fitness. As a matter of fact, a combination of Gi and Ge would be the best possible measure to use. Because of the extra labor involved, however, such a combination is not recommended.

d. Upon the sheet of paper which was used for ranking the teacher's estimates of the pupils' fitness for promotion (see paragraph a), the Gi's of the class should be listed in order from highest to lowest, and arranged so that they will begin on the same line as the names, and run parallel to them. In case there

are two or more identical G_i 's, they should be recorded on separate lines. See Table 18. (If G_e is the measure selected, G_e should be substituted for G_i throughout paragraphs *d* and *e*).

e. The pupil who is ranked highest by the teacher should be assigned the highest G score, that is, the one standing opposite his name. This G score now ceases to be a G_i and becomes a G_t . This G score may not be the G_i score of this pupil. Other pupils receive the G scores opposite their names. These G_t scores should be recorded opposite the pupils' names in the column adjacent to the G_i 's on the Class Record Sheet.

The question may well be asked why we use the G_t score. Although it is not an essential measure, it is useful for many reasons.

a. It permits the incorporation of the teacher's judgment into G_p and hence into the classification of pupils.

b. It permits the teacher to consider other significant factors, such as dependability and emotionality, and to give them weight in the classification of pupils. It should be recognized that the value of the G_t depends upon the ability of the teacher to rate and upon the length of time she has known the pupils.

c. It enables the principal and the teacher to refute parental criticisms of partiality or unfairness. It is often possible to show that G_t is equal to or higher than G_i or G_e and that it thus indicates that the teacher actually rates a pupil higher than objective tests do. If, on the other hand, G_t is seriously below G_i or G_e , it may be well for the teacher to consider this discrepancy carefully, and to inquire into the reasons for it.

d. It gives a measure of how well a teacher knows her pupils. There will not be perfect agreement between G_t and G_e or G_i , since the teacher to some extent is judging traits not measured by the tests. The amount of agreement or disagreement may be judged by inspection. If desired, however, a simple calculation will give a result in quantitative form. For each pupil, the G_e score may be subtracted from the G_t score (G_p may be used instead of G_e). The differences should be recorded as plus or minus, as the case may be. A plus sign indicates that the teacher overestimated the pupil's achievement; a minus sign indicates that she underestimated his achievement. To compare different teachers, average the G_t - G_e differences for each, disregarding signs.

TABLE 18
COMPUTATION OF Gt
GRADE 1L

PUPIL No.	NAMES IN ORDER OF FITNESS	Gt's IN ORDER OF SIZE
20	Harry	2.4
21	Mary	2.2
7	Anna	1.9
8	Dolores	1.5
6	Christine	1.5
18	John	1.5
1	William	1.4
5	Susan	1.3
9	Anne	1.0
19	Joseph	1.0
16	Hildegard	0.9
17	Elizabeth	0.9
3	George B.	0.9
12	Richard	0.9
15	Annabelle	0.8
14	Margaret	0.8
11	Betty	0.7
10	Louis	0.6
4	Gene	0.6
2	Dorothy	0.5
25	George W.	0.5
24	Mary Jane	0.4
23	Mabel	0.3
22	Emily	0.2
13	Fred	0.1

In a study as yet unpublished, Miss Helen Evans has drawn the following tentative conclusions concerning the Gt-Ge relationship:

a. Teachers, on the average, misjudge their pupils by 0.5 of a grade or five months; that is, one-half of a school year.

b. Variation in the size of a class from 30 pupils to 44 pupils does not affect the accuracy of the teacher's estimates.

c. A wide range of ability within a class apparently does not make the teacher's rating any more accurate.

d. Teachers in the lower grades (Grades 3 and 4) tend to rate pupils more accurately than do teachers in the upper grades (Grades 5 and 6).

Step 2. Compute and Record Gp.—The procedure for Step 2 is the same as that for Table 4, except that now we have three

measures to combine for Gp, namely, Gi, Gt, and Ge. The formula then reads as follows:

$$G_p = \frac{wG_i + wG_t + wG_e}{\text{sum of the } w\text{'s}}$$

where w signifies weight. Note that no G_e is available for Grade 1L, and the formula therefore becomes in this case:

$$G_p = \frac{wG_i + wG_t}{\text{sum of the } w\text{'s}}$$

The weights must be decided by the examiner, in the light of the relative worth of G_i , G_t , and G_e . In our sample school, it was decided that they should all have equal weight. The computation of G_p is then simply an average of G_i , G_t , and G_e .

TABLE
CALCULATION OF DIFFERENCES

	GRADE					
	1L	1H	2L	2H	3L	3H
Mean Gp	1.0	1.6	2.2	2.7	3.6	3.9
Norm Gp (Nov. 9)	1.2	1.7	2.2	2.7	3.2	3.7
Differences	-.2	-.1	0	0	.4	.2
	Total of Differences ..					
	Mean Difference					

Step 3. Determine Which Classification Table to Use.—The technique is the same as that used in Table 4. It is repeated here in Table 19 in order to show how to deal with negative differences. The steps in the calculation are as follows: First, the mean Gp scores for all grades should be recorded. Then, below them, the respective norms should be recorded.

In computing the differences, it will be noted that if the mean Gp is higher than the norm, the difference is plus; if lower, it is minus. The sum of the plus differences is 3.3; the sum of the minus differences is -1.3. Adding algebraically, we find that the total of the differences is 2.0. Dividing by 16, we get 0.1 as the mean difference. By referring to Table 6 we find that when the mean difference is between -0.5 and +0.5 we should use the 1.0 Classification Table,—the table designed for

d. In the Gp column, the teacher should find all the scores which are smaller than this classification standard of the half-grade just below. In the Conservative Classification column, the symbol for the half-grade just below should be written opposite these scores. In Table 17, Grade 1L, there are eight pupils, Nos. 2, 10, 13, 14, 22, 23, 24, and 25, whose Gp scores are smaller than 0.7. Opposite these numbers, therefore, we find KH in the Conservative Classification column.

In effect, this procedure does not demote a pupil more than one half-grade, and then only if his Gp is below the classification standard, or the approximate mean, of the grade below.

e. All other pupils remain in their present grades.

CHAPTER XIII

HOW TO HANDLE SPECIAL SITUATIONS

1. The Small School with an Irregular Organization.—In operating any classification system, special situations will develop from time to time. Some of these will be discussed in this chapter.

Many schools which have semi-annual promotion do not have a complete quota of grades. For example, a school might have nine teachers, and one each of the following grades: 1L, 1H, 2L, 2H, 3L, 3H, 4L, 5L, 6L. In reclassifying such a school, it is almost inevitable that pupils will be recommended by the conservative classification for Grades 4H, 5H, and 6H, which are not found in the existing organization.

Let it first be said that the system which has been here outlined should be followed, step by step, up through the conservative classification. By way of illustration, let us suppose that this classification would distribute pupils as follows:

Grade.....	1L	1H	2L	2H	3L	3H	4L	4H	5L	5H	6L	6H
No. of Pupils..	45	35	43	31	40	30	32	20	30	12	17	35

A number of solutions are possible. Some of these are:

a. The principal may secure one, two, or three additional teachers and organize new classes.

b. Combination classes may be organized. For example, we might assign one teacher to Grades 5L and 5H. Or we might organize combinations of Grades 4H and 5L, Grades 5H and 6L. Obviously the 4H and 5L combination would then have 50 pupils. If there is no possibility of transferring pupils to another school, this situation might be met by placing some 5L pupils (whose statistical classification is 5H) in the 5H grade. Some of the 4H pupils could be denied their special promotion and left in 4L.

c. In a city system it is usually possible to transfer pupils to an adjoining school. In this situation, it might be best to transfer the 4H and 6L pupils. We would still have a nine-teacher

school, with the following grades: 1L, 1H, 2L, 2H, 3L, 3H, 4L, Comb. 5L and 5H, 6H.

d. If none of the above proposals seems feasible or desirable, the principal may discard the conservative classification for Grades 4L, 5L, and 6L, and compute the classification according to the rules for a yearly system. It will be recalled that the yearly system presupposes only the low grades, that is, Grades 4L, 5L, and 6L; hence no pupils would be recommended for Grades 4H, 5H, and 6H.

2. A Departmental Organization.—Many schools have a departmental or platoon organization. Difficulties sometimes arise in obtaining teachers' ratings for use in computing the Gt score (grade score according to the teacher's estimate). Often three or more teachers will meet a given class, and there is doubt as to the method of handling ratings.

One type of schedule would assign four classes, for example, two 6L and two 5L classes, to four teachers. Each teacher will then know all the pupils. In obtaining the Gt each teacher should be instructed to treat the two 5L classes as if they were one class. The Gt's of the four teachers may then be averaged. The sixth grade classes may be handled in a similar manner.

Another type of schedule provides for multiple classification. For example, boys will take shop work while girls are taking household arts. In this case, only the teachers who know all the pupils should assist in the ranking. Again, some pupils will go to one teacher for reading, and others to another. No two teachers will know exactly the same group of pupils. In this case it is probably better for one teacher, the home-room teacher for example, to do the ranking. This teacher may consult the other teachers, or the ranking may be done by the group working together.

3. Late Entrants.—Pupils frequently enter school after the initial tests have been given. In order to complete the records many schools make it a practice to give tests once a month to all pupils for whom no records are available. All computations must be made as of the date of the test. For example, the statistical classification will be read in a different column, depending on the number of months the class has been in the half-grade.

4. Ranking Pupils for Gt Early in the School Year.—Many schools administer tests late in the school year, when the

cludes the dullest twenty per cent, or those with I.Q.'s below 90. The Y group includes the middle sixty per cent. The advantage of this plan is that an X group in one school includes about the same type of pupils as an X group in another school. One school may have only one X section, whereas another may have several. A different curriculum should be set up for the different groups, since investigations show certain characteristic differences.

b. When a measure of brightness, such as I.Q., is not available, a crude but convenient method of sectioning is on the basis of chronological age. This may be done in two or more ways. Arrange the pupils' names in each grade in order of chronological age, beginning with the youngest. Then divide the group into the necessary number of sections. If pupils have been properly classified according to the plan described in the foregoing chapters, there will be a general tendency for the younger pupils to be brighter than the older ones. This is because they have reached the same grade level in less time.

Sectioning on the basis of chronological age has several advantages. It is simple and easily handled. Parents will readily recognize that it is reasonable, and there is less likelihood of objection to a pupil's section on the ground that he is deemed to be less bright than a neighbor's child, or that they prefer a particular teacher. Furthermore, pupils of the same chronological age tend to be at approximately the same stage of social development.

6. The School with a Short Term.—Some schools, especially in rural districts, have only a seven- or an eight-months term. Many schools have a nine-months term. This fact does not alter the procedure for classification outlined in the foregoing chapters. Particularly must it not be inferred that all such schools should use the classification table designed for schools which attempt to do 0.9 standard grade per year, since many such schools accomplish a standard grade's work per year.

In a school with a short term there may be confusion about the norms. For example, an eight-months school will under ordinary conditions close on May 1. If tests are administered the latter part of April, the norm or G grade for the fourth grade in a typical ten-months school is 4.8. And yet the pupils in an eight-months school have completed the work of the

grade by May 1, and, strictly speaking, should have attained a norm of 5.0. It would be possible to adjust all norms according to the length of the school term, but this would be inconvenient, and not worth the labor involved. In general, then, the school with a short term should be handled just as any other school; that is, the May 1 grade norm is 4.8.

7. Demotions.—Demotions involve difficulties. At best, they are undesirable. Even when a demotion is indicated according to the conservative classification, it is often best to delay it. Diagnosis and remedial measures will frequently avert a demotion.

If a demotion seems inevitable, the parents should first be interviewed. Above all, the good teacher will aim to preserve the child's feelings. Pupils are persons. An inferiority complex must be avoided. If the parents understand the situation fully, and the facts on which the proposed demotion is based, their coöperation can ease the situation for the child.

It is usually desirable to defer a demotion until an opportune time. For example, if tests are given at the beginning of the year, it is often desirable to delay demotions one month. The teacher's classroom tests will then be available as further evidence of the pupil's achievement. The child often realizes his own weakness. Again, if a child is ill or absent for a time, he has a tangible reason for the demotion. It is often desirable to let a child remain in his grade and receive a non-promotion at the end of the grade, unless he lacks knowledge or skills fundamental to the work of the grade in which he is.

8. Omission of Important Material by Special Promotion.—Pupils who are given special promotions sometimes will skip important units of work. These gaps may be overcome by special assignments or by the help of an adjustment teacher, an opportunity class, or the like. In a large school, a pupil who is given a special promotion may be placed in the average section until he can catch up. It must be borne in mind, however, that we tend to overestimate the importance of these gaps.

9. Computation of Gi from I.Q.—When group intelligence tests are used, it is desirable to give an intelligence test every year. When the individual Binet test is used, the increased accuracy of measurement makes it possible to test pupils at less frequent intervals. At the time of testing, the Gi may be

read directly from the G Table. But the question arises, How shall the Gi be calculated a year later? It must be remembered that the intelligence quotient is relatively constant, whereas the Gi increases from year to year. The problem then becomes that of calculating the Gi, when the I.Q. is given. The procedure is as follows:

Step 1.—The mental age, as of the desired date, should be calculated by means of the formula—

Mental age = chronological age \times intelligence quotient. In this calculation, the intelligence quotient is regarded as a decimal. Chronological age is expressed in years and decimals of a year.

Step 2.—In Table 21, the Gi corresponding to the estimated mental age may be read.

Illustration: Assume that at age 11 years a pupil's I.Q. was found to be 90. At the date for which the Gi is desired, assume his age to be 12 years 5 months. The procedure is as follows:

12 years 5 months = 12.4 years

Future mental age = $.90 \times 12.4 = 11.2$. According to Table 21, 11.2 years converts to a Gi of 5.4.

For a fuller treatment of the classification of pupils in the high school the reader may consult the following reference:

Symonds, Percival M., *Measurement in Secondary Education*, The Macmillan Company, New York, 1927.

BOOK FOUR

PROGRAM OF MEASUREMENT FOR
PROGRESSIVE SCHOOLS

CHAPTER XIV

THE COMPREHENSIVE TESTS¹

The series of four tests described in this chapter was made for the purpose of stimulating achievement in education, of fairly evaluating achievement, and of changing the idea of achievement still further from exclusive concern with subject matter toward interest in abundant democratic living.

The *Intelligence Test* was constructed in order to measure intelligence; the *Educational Background Questionnaire* to measure aspects of educability not measured by intelligence tests; the *Comprehensive Achievement Test* to measure the widest possible range of desirable learnings and to motivate growth; and the *School Practices Questionnaire* to identify the curriculum of democratic activity, to evaluate school practices, and to motivate growth.

These four tests are used for illustration because they have just been completed, because they were developed as a unit, because no other group of tests specially designed to fit together are so truly comprehensive, because they form a unit particularly acceptable to progressive schools, and because the existence of descriptions of them and directions for using them saved the author much time in writing Chapters XIV and XV. Other intelligence tests, as good or better than this one, may be substituted for it. At the present writing, there are no satisfactory substitutes for the other three. Most of the skill aspects of the *Comprehensive Achievement Test* are more satisfactorily measured by, say, the Iowa, Metropolitan, Modern School, or New Stanford Achievement batteries.

1. THE INTELLIGENCE TEST

The *Intelligence Test* is a multiple-choice test of the mental type, used from the third through the ninth grade.

These principles have guided the construction of the foregoing intelligence test.

¹ Quoted or adapted in part from *A Comprehensive Test Program-Manual for Teachers* by William A. McCall and John P. Herring, and with the kind permission of Laidlaw Bros., Chicago.

1. *An intelligence test should be a learning test which extends backward rather than forward.*—These four—the number of desirable neural connections, the organization of these connections, the ease of forming and breaking connections, and the permanence of connections—are the chief characteristics of an individual which a test must measure if it is to be a good intelligence test. Two methods have been proposed for testing these preëminently valuable characteristics. One method is to confront a pupil with a learning situation which varies from very simple to very complex. A measurement of the number of points learned, the maximum complexity of the thing that could be learned, the rate of learning, and the persistence of the things learned, would give a measure of the pupil's four prime characteristics. The inherent difficulties in conducting such learning tests are so great that another testing method is in almost exclusive use. This method takes samplings from the abilities which a pupil has, during his whole life, succeeded in developing. While this is also a learning test method, the learning test extends all the way from the present back to birth rather than from the present to a brief future.

No great growth of intelligence could occur if each night of sleep wiped out the learning of the day as each night in Valhalla healed the wounds of the day's battle. There is a popular belief, fallaciously transferred from bank accounts to individuals' memories that "easy comes, easy goes," and hence superior intelligence cannot possess both superior plasticity and superior permanence. Not only "to him that hath shall be given" but to him that hath *has been* given. For the adage that he who learns quickly forgets quickly is not based on facts but upon a sympathetic desire to comfort stupid folks.

While the test presented here is a learning test it does not measure traits which are the direct object of school instruction. In fact it has been demonstrated that direct teaching of the meaning of all the words in a similar test containing more difficult words contributed little to a pupil's score.

This test to an unusual degree is reserved for the measurement of intelligence itself by being freed from the measurement of achievement. All the words in the test are found in Gates's easiest thousand words in his *Reading Vocabulary for Primary Grades*, which is based upon Thorndike's count of four

and one-half million words and upon other studies. In Grade IV and above, therefore, the test is not a test of ability to read words, and it is obviously not a test of ability to read sentences or paragraphs. The mechanics of the test, being simple to learn, are not open to the charge of allowing an element of achievement to enter the score.

2. *An intelligence test should measure the largest possible number of traits.*—While it is probably true that every mental trait or set of neural connections boasts no aristocratic exclusiveness, but exemplifies even Nature's predilection for democracy by partially combining with other traits to constitute a coördinating neural hierarchy, nevertheless, every trait retains a portion of its individuality or exclusiveness! For this reason, the larger the number of traits measured, the safer the diagnosis. A test which measured but a few traits might happen to strike just those mental functions in which the pupil, for some accidental reason, was specially strong or specially wanting. The assayer takes many samples from many points in the ore bed. One of psychology's important criteria of superior or inferior intelligence is the differences in the ability for minute analysis, "piecemeal activity," or to deal with subtle elements of a situation. In neural terms this means that the more intelligent individual has more neural connections for any one situation. To a stupid individual a ripe peach will probably suggest gastronomic satisfaction only, while to the more intelligent it suggests this to be sure, but it may also suggest the flush of dawn, the blush of a maid, the softness of a baby's cheek, or the fruit of the Tree of Knowledge! The flower in the crannied wall was more to Tennyson than a pretty weed to adorn a vain buttonhole! To those with numerous neural connections "every chip sprouts wings to bear a god" and falling apples cause a flow of ideas as well as a flow of saliva! To a Woodberry, "a rose shadows us with Persia, or a single lotus blossom unbosoms all the Nile."

This test, containing some thirty different types of items, measures a variety of kinds of intelligence so that the student must persistently readjust his mental set by seeking to divine thirty or more different principles of relationship among ideas. That such mental flexibility is called for is not often noticed when the test is first seen, because the items look so much alike.

Persistence in the face of bafflement is probably also a factor in the score, as it is in the effectiveness of intelligence generally.

3. *An intelligence test should measure samplings from the relatively more differentiating traits.*—The ideal way is to measure every trait that contributes to intelligence, attach weights to the various traits according to the amount of their contribution to intelligence, and add. The resulting sum would be a perfect measure of intelligence. Even if we knew how to test every trait and just what weights to attach to each, time would compel us to confine our attention to the more differentiating traits.

But how may we know what are the differentiating traits? Let us proceed by the process of elimination. We can eliminate those traits in which man is little or not at all superior to the animals. Certain elemental functions such as keenness of vision, hearing, and smell, speed of simple muscular responses like running and tapping and the excellence of the neural functioning in connection with breathing, digesting, and other organic functions, all these are of great importance. There are more failures from indigestion than this world dreams of. After a certain minimum these traits have little differentiating value. In them the brute and the stupid human are about the same as the genius. The intelligence tester will do well to steer clear of simple sensorimotor tests and seek out those traits which chiefly distinguish man from the brute, and genius from stupidity. Simple observation of these distinctions will point the way toward differentiating traits.

Both observations and correlations with semi-satisfactory estimates of intelligence have indicated that intelligence tests should measure such traits as the ability to analyze a complicated situation, to attend to many elements at one time, to easily and effectively shift from one mental set to another, to deal with abstract symbols and relationships and the like.

By compelling examinees to practice repeated divination or insight, the test becomes more valid as a measure of intelligence. The high validity or reliability of this type of test has been attested by the studies of McCall and Speer on the elementary levels and of Trabue on the maturer levels. Abelson and Bar-

thelness have each devoted a doctor's dissertation to the type of test item used.

4. *An intelligence test should measure only those traits which every pupil has an equal opportunity to develop.*—This means that the test material and methods of the test should be drawn from the social medium common to all children. Theoretically there should be *equal* opportunity to learn the test material, but practically about all that can be provided for is *ample* opportunity. Those traits should be measured which are least influenced by such differential agencies as school *vs.* non-school training, city *vs.* rural life, masculinity *vs.* femininity, luxury *vs.* poverty, etc. A country boy might easily show up unfavorably in comparison with his city cousin in reacting to questions about elevators, skyscrapers, subways and rollerskates, while the situation might be reversed if the questions dealt with hay-mows, disc-harrows, silos, dibbles, copperheads, and yellow jackets.

This test is drawn from a medium common to all school children in America—very simple words.

5. *An intelligence test should show a higher per cent of correct responses with each increase in chronological age.*—This principle holds up to the age when intelligence matures, which is supposed to be not far from 20 years of age. The fundamental assumptions underlying the intelligence test as customarily used are that the total amount of knowledge, skill and power acquired by an individual (a) is a measure of his present intelligence, (b) is proportional to his inherited intelligence, and (c) is prophetic of his future intelligence. These three points mean that if one infant has a native endowment twice that of another child, he will develop proportionately faster until intelligence matures and hence will at every stage of his life be proportionately superior to the originally inferior individual.

6. *An intelligence test should measure the ability to transfer training.*—One of the great advantages of possessing numerous neural connections which are effectively organized is that they guarantee wide-scale transfer. The genius makes everything grist which comes to his mill. He can transfer both Latin and Algebra to just about anything. The stupid individual, on the contrary, lacks this nimbleness of wit. He can be trained but can be educated only with difficulty. He would make a fairly good showing if the test contained material upon which he had

had direct training. When the test presents tasks for which he has had no specific training his existing neural connections are unable to deal with the new situations. This difference between individuals in their power to deal with situations for which they have had no specific training is so significant and marked that intelligence might well be defined as the power to transfer training. An intelligence test which does not measure this ability is certainly imperfect.

7. *An intelligence test should measure over a wide range of intelligence.*—This is not an absolute requirement but it is important in a test that is designed as this one, to serve as a basis for standardizing school marks throughout the nation. By incorporating the divination feature, this test was made suitable for a phenomenal range of ability, namely from Grade III through the university. It might almost be said that the test increases in psychological difficulty with progress through the grades, for maturer persons perceiving subtler relationships, solve the same items on a more difficult level.

INTELLIGENCE TEST¹

FOR GRADES 3 THROUGH 9—FORM I

Number Right_____ G Score_____ MA_____ IQ_____

Name_____ Sex_____ Grade_____

Age: Yrs._____ Mos._____ Date_____ Teacher_____

School_____ City_____ State_____

Instructions. Write your name, grade, and so forth in the blanks above.

Look below at the first set of words: *hat coat tree dress shoes*. The word *tree* does not belong with the others. Is that right? The word *tree* is the third (3) word, so 3 is written in the space at the right.

Look at the next set of words: *chair yes dog no pup*. The word *chair* does not belong with the others because *yes* and *no* belong together and *dog* and *pup* belong together. Is that right? Since *chair* is the first (1) word, 1 is placed in the space at the right.

Look at the set of numbers. The fifth number, 19, does not belong with the others, so 5 is written in the space at the right. Is that right?

Look at the fourth set of words. The word *pie* does not belong with the others because these others make a sentence: *This test is fun*. Is that right? Since *pie* is the second word, 2 is placed in the space at the right.

¹ Published by Laidlaw Brothers, Copyright 1937 by Laidlaw Brothers, Inc.

Look at the next set of words. Put the right number in the space.

Look at the last set of words. Put the right number in the space.

1. hat	coat	tree	dress	shoes	3
2. chair	yes	dog	no	pup	1
3. 12	14	15	13	19	5
4. test	pie	fun	is	this	2
5. hard	long	soft	eye	short	_____
6. star	coat	sheep	wool	cloth	_____

Look at me when you finish (*pause*). What is the answer to Item 5? Why? To Item 6? Why? Do you understand?

On the following pages, do every item. Take them in order. Do not skip. *Do not waste time on a hard item*; do it the best you can quickly and go on to the next one. You will have plenty of time if you do not waste it. Your score will be the number you do correctly in the time allowed. When you are told to do so, turn this page and begin. You will be stopped exactly 40 minutes later.

TO THE EXAMINER: Read instructions aloud while the children read silently. Give no help after the test begins, beyond seeing that instructions are followed.

1. boy	dog	cat	doll	rat	_____
2. book	eat	read	bread	sky	_____
3. girl	boy	them	she	him	_____
4. eye	foot	nose	chin	lip	_____
5. party	red	to	come	my	_____
6. cold	white	black	snow	hot	_____
7. seed	tree	root	leaf	sun	_____
8. 5	4	2	8	6	_____
9. boy	dog	play	moon	bark	_____
10. wood	coal	rock	fire	burn	_____
11. 1	2	7	3	5	_____
12. my	yours	hers	your	mine	_____
13. she	I	is	teacher	the	_____
14. good	bad	clean	pretty	dirty	_____
15. rock	mud	dirt	sand	air	_____

Go to the next page.

16. star	sun	earth	sea	moon	_____
17. him	give	that	ran	book	_____
18. 2	10	8	14	6	_____
19. am	we	he	she	will	_____
20. toy	he	town	to	went	_____
21. train	arm	city	car	hand	_____
22. man	child	toy	horse	cow	_____
23. boys	come	pie	like	all	_____
24. big	slow	small	run	quick	_____
25. egg	tree	bird	apple	lily	_____
26. like	we	bad	food	good	_____
27. carrot	apple	orange	banana	pear	_____
28. 3	6	14	12	9	_____
29. gray	red	yellow	green	blue	_____
30. goose	crow	hen	corn	duck	_____
31. door	floor	wall	bed	window	_____
32. is	fun	this	these	game	_____
33. talk	eye	ear	mouth	hear	_____
34. 5	16	11	6	8	_____
35. ear	pen	nose	pencil	paper	_____
36. finger	leg	arm	toe	ear	_____
37. yes	do	no	not	don't	_____
38. smile	cry	frown	laugh	sneeze	_____
39. horse	wagon	cow	sheep	pig	_____
40. hour	year	day	month	night	_____
41. the	black	sun	table	is	_____
42. wall	floor	house	city	door	_____
43. slow	fast	late	hard	soft	_____
44. knife	saucer	fork	dish	cup	_____
45. 11	16	4	8	2	_____
46. she	mother	he	is	the	_____
47. since	all	many	every	much	_____
48. pony	kitten	sheep	cat	lamb	_____
49. horse	cat	runs	fish	climbs	_____
50. hen	calf	egg	chick	cow	_____
51. shoes	hat	coat	dress	car	_____
52. 3	11	6	66	33	_____
53. fruit	cow	sun	tree	calf	_____
54. fur	tail	eye	wheel	ear	_____
55. has	his	hers	ours	its	_____

Go to the next page.

56. rats	dogs	cats	run	mice	_____
57. fly	ball	top	throw	spin	_____
58. bad	good	kind	true	brave	_____
59. 1	16	13	11	6	_____
60. he	to	of	her	him	_____
61. lion	frog	tiger	dog	cow	_____
62. feed	food	eat	fed	ate	_____
63. go	get	give	come	send	_____
64. tag	ball	easy	play	game	_____
65. pen	hand	see	touch	eye	_____
66. foot	leg	shoe	nail	toe	_____
67. both	many	all	some	each	_____
68. 20	6	15	18	4	_____
69. run	skip	hop	walk	crawl	_____
70. book	me	story	a	tell	_____
71. me	she	my	hers	you	_____
72. call	loud	sing	talk	speak	_____
73. fish	bird	crawl	hop	snake	_____
74. which	what	that	when	this	_____
75. she	boy	the	is	he	_____
76. pond	lake	rain	river	ocean	_____
77. up	from	down	to	around	_____
78. sheep	milk	horse	cow	wool	_____
79. water	ate	drank	egg	swim	_____
80. 21	3	18	9	15	_____
81. fire	water	flame	smoke	heat	_____
82. right	in	came	out	left	_____
83. how	when	where	what	are	_____
84. hat	cloth	dress	glove	shoe	_____
85. please	tell	me	go	with	_____
86. wrong	left	top	right	bottom	_____
87. hammer	saw	nail	see	ax	_____
88. woman	man	for	work	very	_____
89. 21	18	16	7	6	_____
90. apple	peach	pear	beet	plum	_____
91. see	sight	sat	saw	sit	_____
92. 4	8	4	1	3	_____
93. yes	the	he	boy	is	_____
94. will	could	can't	can	won't	_____
95. dog	is	our	teeth	white	_____

Go to the next page.

96. fat	slow	old	gray	fresh	_____
97. which	why	high	down	under	_____
98. cold	soup	burn	mouth	hot	_____
99. 3	13	8	15	18	_____
100. worse	when	better	wrong	good	_____
101. frost	ice	snow	hail	dew	_____
102. we	fun	like	hit	good	_____
103. is	were	are	be	was	_____
104. they	man	it	he	book	_____
105. foot	head	shoe	coat	hat	_____
106. 6	9	11	15	18	_____
107. and	it	or	but	if	_____
108. bread	sheep	fruit	meat	tree	_____
109. gone	start	go	stop	come	_____
110. this	them	those	that	these	_____
111. light	still	sound	dark	loud	_____
112. get	give	got	gave	go	_____
113. his	who	what	name	is	_____
114. 2	16	9	7	17	_____
115. bark	stars	tree	leaf	nuts	_____
116. above	off	under	on	in	_____
117. as	cloud	blue	such	green	_____
118. come	go	stay	hop	run	_____
119. too	that	a	also	an	_____
120. 18	10	13	11	16	_____
121. you	kiss	will	no	me	_____
122. we	can	you	they	is	_____
123. for	from	here	there	to	_____
124. help	we	them	chair	can	_____
125. yes	on	own	but	want	_____
126. Jim	Jane	Helen	Ruth	Bess	_____
127. 19	5	4	25	16	_____
128. apple	grape	plum	cherry	pear	_____
129. above	over	around	under	below	_____
130. money	play	rest	work	fun	_____
131. store	chalk	mother	teacher	child	_____
132. little	big	tall	large	small	_____
133. sleep	skate	run	swim	jump	_____
134. warm	dry	hot	wet	cold	_____
135. milk	ice	store	water	cream	_____

Go to the next page.

136. teeth	jaws	ear	mouth	lips	_____
137. 29	27	14	30	23	_____
138. little	slow	easy	short	same	_____
139. big	ever	never	little	often	_____
140. game	ball	foot	coat	bat	_____
141. give	shoe	me	gave	your	_____
142. head	hand	hair	arm	finger	_____
143. ship	wave	sail	sea	cloud	_____
144. 19	9	14	8	11	_____
145. very	but	some	more	any	_____
146. moon	sun	night	dark	day	_____
147. fast	noisy	still	slow	moving	_____
148. sun	light	man	tree	girl	_____
149. where	why	down	above	below	_____
150. very	am	he	I	bright	_____

Use all the time to improve your answers.

A knowledge of a pupil's I.Q. should be of very great value to any teacher of any subject, for the size of a pupil's I.Q. is an index of his general mental brightness or mental alertness. As Terman points out the most important fact about a pupil, next to character, is his I.Q. The significance of I.Q.'s of varying sizes is brought out below:

Above 140	Genius or near genius.
120-140	Very superior intelligence.
110-120	Superior intelligence.
90-110	Normal or average intelligence.
80-90	Dullness.
70-80	Borderline deficiency, sometimes feeble-mindedness.
Below 70	Definitely feeble-minded.

These determinations of mental age and intelligence quotient not only furnish valuable teaching guides but also provide the basis for educational guidance through a knowledge of a pupil's capacity to profit by general education and pursue particular subjects.

One problem in education is to locate the educational objectives. Another is to locate somebody who has the capacity to attain these objectives—to find somebody who is educable. Pigs, sheep, cows, horses, dogs, and other domesticated animals have widely varying capacities to learn. While the percentage of illiteracy is high, these animals have a more or less definite cur-

riculum and are taught certain lessons by their owners. It is only human beings, however, who are considered to have enough capacity to learn to make systematic, prolonged education profitable.

But the technique of diagnosing capacity to learn does not end with the classification of an animal as a human animal. The range of capacity to learn among humans is greater than the difference between humans in general and dogs in general. The overlapping of capacity is so great that a considerable per cent of humans have a capacity to learn which is inferior to the geniuses among dogs, cats, monkeys, and other much reviled creatures.

The first measurements of capacity to learn were simple unstandardized observation of children by parents and neighbors. These measurements were inevitably inaccurate because of numerous constant errors such as parental vanity, neighborly jealousies, absence of constant or fair standards of estimate as well as other more subtle errors. These subjective measurements were probably more accurate a generation ago than at present, because numerous progeny facilitated the development of surer standards of measurement.

As a result of parental measurements the extremely dull children were kept at home or sent to special institutions, while those with a greater capacity were sent to school.

When children are dumped into the hopper of the educational mill, they enter a great and more accurate selective machine. Every stage of education from the kindergarten through the university is engaged in the process of selection. In a very real sense our schools are as much selective as educative agencies. Every teacher takes her toll, though gallantry forbids saying that this is a case of the devil takes the hindmost. There is a miller whose water mill on the Tennessee River grinds the corn for the farmers for miles around. The miller always takes his toll from the best bag of corn. Teachers are more generous; they take toll from the children whose capacity to learn is least. Each year the ranks of this grand army of children grows thinner and thinner. The Ph.D. or the equivalent is the educator's reward for the students who have been able enough or clever enough to escape the clutch of all the teachers.

More and more educational selection is becoming an important function of the school. Children are being committed to institutions for the feeble-minded. This is frequently construed as

a stigma upon both children and parents. Private schools deny entrance to children whose learning capacity is judged to be below a certain standard. Public schools are sending pupils to special classes for the mentally slow. Dull pupils are denied promotion. Certain public schools group pupils within each grade according to learning capacity. Other public schools refuse admission to any whose learning capacity is not unusually great. Some countries, recognizing that their greatest asset is their children of genius and that these geniuses belong to the community rather than to particular parents, are selecting these children for a special education.

When matters of such critical importance to the individual are at stake a democracy will not long tolerate a system of educational selection which does not utilize the most thoroughly scientific, impartial, impersonal, and rigidly standardized technique possible. Standardized educational and psychological tests, inaccurate though they may be, are rapidly becoming recognized as the best means for educational selection. It is but a question of time until they supplant the traditional selective mechanism of home and school.

Psychologists are now able to tell with considerable accuracy whether a child possesses an I.Q. which will ever make it possible for him to do the work of a particular school or institution or grade in a school. Further, they are able to determine whether a child's mental age is now sufficient to learn the work of a particular grade. Terman's experience leads him to the conclusion that the 60 I.Q. pupil will not be able to do work beyond Grade III or IV. The 70 I.Q. child will not be able to do work beyond Grade V or VI. The 80 I.Q. will reach his limit about Grade VII. The 90 I.Q. pupil may by dint of much persistence go through high school. E.Q.'s of 60, 70, 80, and 90 for pupils whose educational opportunities have been normal may be interpreted like similar I.Q.'s. Even the attainment listed above cannot be reached until the mental age or educational age has sufficiently developed and this means considerable chronological retardation.

Since social judgment is the final criterion of intelligence, why not employ it exclusively? Tests are resorted to not only because they are far more economical but particularly because they are impersonal and prophetic. History has changed too many pilories to monuments, and parental evaluations of children have

been too frequently reversed for us not to know that subjective judgment often tends to be prejudiced unless the observations are safeguarded with unusual care. For this reason the relatively ice-cold mob-proof, carefully validated intelligence tests are coming to be used more and more for intellectual determination. Intelligence tests possess the further incalculable value of being prophetic. Tests do not wait until success is achieved before passing verdict. They do not lay roses on the grave of a genius, but crown him in childhood.

For more information of a popular nature about the measurement of intelligence, the reader may consult these references:

Freeman, Frank N., *Mental Tests*, Houghton Mifflin Co., New York.

Pintner, Rudolf, *Intelligence Testing* (New Edition) Henry Holt and Co., New York, 1931.

For a thorough technical discussion, he may consult this reference:

Thorndike, E. L., *The Measurement of Intelligence*, Bureau of Publications, Teachers College, Columbia University, New York.

2. THE EDUCATIONAL BACKGROUND QUESTIONNAIRE

The *Educational Background Questionnaire*, a multiple-choice questionnaire for Grades IV through IX,¹ supplements intelligence tests as measures of educability. If education were still almost exclusively interested in intellectual achievement, the intelligence tests would serve the purpose. But the school curriculum is now being gradually widened to include experiences (such as learning to work harmoniously with people of different ages) which depend upon good homes and good neighbors more than upon intelligence. That is why intelligence tests need to be supplemented for the purpose of measuring the total educability of the child. An important assumption, therefore, underlies this questionnaire, namely, that ability to learn depends not only upon native intelligence but also upon factors in part cultural. Such factors are, for example, health, social and economic status, schooling of the family and of the community, facilities for education in the home and in the community outside of the school, and the educational pressure of family and community. The

¹ But it is usable also in high school and college.

function of each item in the questionnaire is to help accumulate the evidence which intelligence tests do not provide as to the degree of educability of the child. That purpose makes the test unique. An intelligence test plus a background test of this kind should measure educability more completely than either one of them alone. There can be no doubt that both heredity and the non-school environment play a role in making the child ready to avail himself of what his school has to offer him.

The test also provides information about the home and the community to assist the teacher in her adjustment to the child's needs. The daily counseling and guidance of the young by means of every knowledge of personality and environment is near the heart of education. Indeed, the modern teaching process has become in large part a series of acts of both individual and group counseling about every aspect of social, physical, and material relationships. The test will help show why a child behaves as he does in school and will suggest both purposes for home contacts and remedial measures.

The modern teacher combines in one person (and therefore in one process) something of sociologist, psychologist, psychiatrist, mental hygienist, counselor, parent, friend, philosopher. He does not specialize in any one of the fields, but in a combination of them. The questionnaire plays directly into his hands by providing information about home and community.

A third use of the questionnaire, important like the others but more difficult, is to educate the community. The statistical summaries for a school give a picture of the health, attractiveness, culture, wealth, education, and enterprisingness of the community itself. The averages tell something, even, of the educability of the community. They are a starting and continuing point for acquainting the school with the life around it. They hold out the possibility of the long-time planning and developing of community life in conjunction with school activity, of old and young working side by side, of school and the rest of the community learning in one coherent social whole the democratic ways of life.

Year by year the patrons, going oftener to their school, and the students oftener to their community, disclosing common problems in common counsel, together facing situations, together experimenting, together discussing values, will increase the scores in the questionnaire. Higher scores imply enlarged

opportunities for improvement in the achievement test, and *vice versa*. The two tests, dynamically used, act reciprocally each to raise the score of the other. There ought to be persons from the community spending an increasing amount of time, money, and effort collaborating with the school. There ought to be a democratic give and take of initiative, planning, criticism, experiment, and enjoyed outcomes. The number of participants ought ever to be on the increase, and the benefits ought to be sufficient to guarantee to each his continuance and his enthusiastic commendation of such concourse of diverse persons and groups. Although the influence of the community upon the education of the child has usually been a fixed asset or liability, it ought to be not fixed but dynamic. School and community ought to influence each other in continuous, joint growth.

What to do for a beginning is especially suggested in the *School Practices Questionnaire*, which is a means of appraising that curriculum which the school makes actual for the child, and it will be further suggested to those who use the whole series of tests and follow the implications which will emerge from so doing. The tests have been so made that paths not at first evident will open before such persons.

Since the test is new in type and what it will do is not necessarily evident to those who read it for the first time, it is wise to foresee certain possible misunderstandings.

Some may think that the reading difficulty is too great. The language has been painstakingly scrutinized in the light of Thorndike's and Gates's studies of vocabularies. It has been repeatedly revised after trials with the lower levels of low fourth grades. A few words, like *spinal meningitis*, which there is no way to avoid, are likely to be read by children who have had the experiences represented by the words. Teachers, moreover, are directed to help students with word meanings throughout the test.

Some may think that orphans are penalized by the test. They are, but no more probably than they are by life itself, because they do not have the same chance as other children have to develop satisfactory traits, such as an adequate feeling of security in personal relationships.

Some may think that the only child is penalized. He is, but no more probably than in life. He is not likely to learn certain social relationships, like getting along with people, so readily as he

could in a home with other children of ages suited to his needs. His social educability is more likely to be low.

Some may think that a child with a reasonable number of brothers and sisters is not necessarily made more receptive to social education. He probably is if his mother is adequate, and he probably is not if she is inadequate. Question 82, *How many brothers have you who live with you in your home?* needs, therefore, to be supplemented by Question 111, *Does your mother pay little attention to what you do or does she help you with the important things you do or does she direct everything you do?* Question 82 and many others are thus modified, sometimes by means of several questions each, scattered in different parts of the test, until, through this method of supplementary items, a sufficiently qualified total score is secured.

NUMBER RIGHT											
PART	A	B	C	D	E	F	G	H	I	J	TOTAL

EDUCATIONAL BACKGROUND QUESTIONNAIRE ¹

FOR GRADES 4 THROUGH 9

Name _____ Sex _____ Grade _____
 Age: Yrs. _____ Mos. _____ Date _____ Teacher _____
 School _____ City _____ State _____

Instructions. Write your name, sex, grade, and so forth in the blanks above.

On the following pages are many interesting questions. Your answers to them will help us to make your school life happier.

Read Question 1 below. The answer, 4, is circled in the column to the right of the question.

Read the rest of the questions and circle the true answers yourself.

- | | |
|---------------------------------|---|
| 1. How much are two and two? | 3
④ |
| 2. How often do you have colds? | always
usually
often
seldom
never |

Go on to the next page.

¹ Published by Laidlaw Brothers. Copyright 1937 by Laidlaw Brothers, Inc.

-
3. How many of your close friends have finished the eighth grade, both 8A and 8B? most
half
a few
none
4. Which child are you most like? *John* usually plays by himself. *Rose* usually plays with other children. John
Rose
Ned
5. Which one is best to do? a
Take a (a) cold shower every day (b) hot bath b
every week (c) warm soap bath almost every day c
d
6. Circle each one that you have at your home. 0
automobile 1
running water 2 or more
electric lights
How many did you circle?
7. Do you know how to mark the best answers? yes
no
not sure

Look at me when you finish (*pause*). How did you answer Question 2? 3? 4? 5? 6? 7?

On the following pages answer *every* question. Take the questions in order. Do not leave out any. Whenever you cannot answer some question, ask the teacher to help you. When you are told to do so, turn this page and begin.

TO THE EXAMINER: Help the children with the practice questions on the front page. See that every child answers every question in the questionnaire. Help children to answer the questions whenever they need help throughout the questionnaire. Do not influence the children for or against any answer; merely help them to give the truest answers. Do not let any child spend much time on any one question. Children will usually finish in 40 minutes, but there is no time limit.

A. YOUR HEALTH

1. How much of the time are you well? most
half
seldom
2. Do you feel tired? usually
often
seldom

Go on to the next page.

-
- | | |
|--|------------------------------------|
| 3. Is it hard for you to go to sleep? | usually
often
seldom |
| 4. About how many hours do you sleep each day and night? | 6 or 7
8 or more |
| 5. Do you feel rested when you wake up? | usually
often
seldom |
| 6. Do your teeth ache? | usually
often
seldom |
| 7. Does your stomach ache? | usually
seldom or
never |
| 8. Do your eyes hurt you or trouble you and make it hard to read, or to see the blackboard, or to study? | always
often
seldom
never |
| 9. Circle each of these sicknesses that you have had. | |
| chicken pox | |
| smallpox | |
| measles | |
| scarlet fever | |
| typhoid fever | |
| spinal meningitis | |
| pneumonia | 0 |
| infantile paralysis | 1 |
| diphtheria | 2 |
| How many did you circle? | 3 or more |
| 10. How often do you have bad colds? | often
seldom
never |

B. HEALTH OF THE FAMILY

- | | |
|---|----------------------------|
| 11. Is there sickness in your home? | usually
often
seldom |
| 12. Is anyone in your home usually or always sick in bed? | yes
no |

Go on to the next page.

-
13. How many times in the last twelve months
has a doctor come to your home to cure grown
persons who were sick? 3 or less
4 or more
14. How many children in your family have died? 0
1 or more
15. Are your *own* parents living? both
father only
mother only
neither
16. How many of your grandparents are living? 0
1
2
3 or 4

C. YOUR SCHOOL PLANS

17. Have you passed the eighth grade, both 8A
and 8B? yes
no
18. Shall you go to school until you pass the eighth
grade? passed already
yes
no
not sure
19. Shall you go to school until you finish senior
high school? yes
no
not sure
20. Shall you go to college? yes
no
not sure
21. Shall you stop going to school as soon as you
can? yes
no
22. How many schools have you gone to, in the last
three years? 1 or 2
3 or more
23. In how many countries besides the United
States have you been, in the last three years? 0
1
2
3 or more
24. In how many states of the United States have
you been, in the last three years? 1
2
3
4 or more

Go on to the next page.

D. SCHOOLING OF YOUR FAMILY

25. Do (Did) your own parents read and write English? both
 father only
 mother only
 neither
26. Did your own parents finish the eighth grade? both
 father only
 mother only
 neither
27. In answering this question and other questions like it, count the schooling received in this and other countries. Did your own parents go to high school? both
 father only
 mother only
 neither
28. Did your own parents go to college? both
 father only
 mother only
 neither
29. Did your own parents go to business or vocational school? both
 father only
 mother only
 neither
30. How many of your grown relatives, not father and mother, went to high school? all
 most
 half
 a few
31. Have you any brothers or sisters who went to college? yes
 no
32. Have you any brothers and sisters who *will* go to college? yes
 no

E. SCHOOLING OF YOUR COMMUNITY

33. How many of the grown persons that you know best went to high school? all
 most
 some
 none
34. How many of the grown persons that you know best went to college? most
 half
 a few
 none

Go on to the next page.

-
- | | |
|---|------------------------------|
| 35. How many close friends have you who will go to school until they finish senior high school? | 0
1
2
3 or more |
| 36. How many close friends have you who will go to college? | 0
1
2
3 or more |
| 37. How many close friends have you who will leave school as soon as they can? | 0
1
2
3 or more |
| 38. How many of the grown persons that you know best read and write English easily? | all
most
half
a few |

F. YOUR STUDY HABITS AND CONDITIONS

- | | |
|---|--------------------------------|
| 39. Is there a room in your home where you can study by yourself? | yes
no |
| 40. Are you interrupted or disturbed when you study at home? | always
often
seldom |
| 41. Which child are you most like?
<i>Sue</i> studies when she feels like it.
<i>Ned</i> studies regularly every day.
<i>John</i> studies whenever he needs to, no matter how long it takes. | Sue
Ned
Roxy
John |
| 42. Have you a good place to keep your books and papers? | yes
not good
none |
| 43. Do you study an hour or more each day at home? | yes
no |
| 44. About how many books are there in your home? | 99 or less
100 or more |
| 45. How often is it quiet around you inside your home when you study? | usually
sometimes
seldom |

Go on to the next page.

-
46. How noisy is it just outside your home? very noisy
noisy
quiet
47. How many daily papers does your family take regularly? none
1 or more
48. How many magazines do you and your family take regularly? none
1 or more
49. Do you like to work at hard puzzles? yes
no
50. Is there room for outdoor games in your yard? yes
no
51. How many persons older than you often keep you from having a good time when you are playing? 0
1
2
3 or more
52. About how many books have you at home, not schoolbooks, that belong to you yourself? 19 or less
20 or more
53. About how many hours do you work anywhere for pay from Monday morning to Friday night? 0 to 5
6 or more
54. About how many evenings each week are you at home the whole evening? 0 to 4
5 or more

G. SOCIAL-ECONOMIC STATUS OF
YOUR FAMILY

55. Have you a telephone in your home? yes
no
56. Circle each one that you have in your home.
vacuum cleaner
electric clock
electric iron
How many did you circle? 0
1
2 or more
57. Circle each one that you have at your home.
automobile
paid servant
bathroom for your family alone
How many did you circle? 0
1
2 or more

Go on to the next page.

58. Circle each one that you have in your home.
- | | |
|---------------------------------|-----------|
| electric ironing machine | 0 |
| electric lights | 1 |
| gas lights | 2 |
| power washing machine | 3 or more |
| <i>How many did you circle?</i> | |
59. Circle each one that you have in your home.
- | | |
|---------------------------------|-----------|
| organ | 0 |
| piano | 1 |
| radio | 2 |
| phonograph | 3 or more |
| <i>How many did you circle?</i> | |
60. Circle each one that you have in your home.
- | | |
|---------------------------------|---|
| cold running water | 0 |
| hot running water | 1 |
| <i>How many did you circle?</i> | 2 |
61. Circle each one that you have in your home.
- | | |
|---------------------------------|-----------|
| furnace or oil burner | 0 |
| gas range | 1 |
| electric range | 2 or more |
| electric refrigerator | |
| gas refrigerator | |
| <i>How many did you circle?</i> | |
62. Does your home have more rooms or more persons in it?
- | |
|--------------|
| more rooms |
| more persons |
| same number |
63. About how many times a year do you go to a dentist?
- | |
|-----------|
| 0 |
| 1 |
| 2 or more |
64. About how many times a week do you go without breakfast?
- | |
|-----------|
| 0 |
| 1 |
| 2 or more |
65. How many other persons sleep in your bedroom?
- | |
|-----------|
| 0 |
| 1 |
| 2 or more |
66. How many other persons sleep in the same bed with you?
- | |
|-----------|
| 0 |
| 1 |
| 2 or more |
67. How many windows has your bedroom?
- | |
|-----------|
| 0 |
| 1 or more |

Go on to the next page.

Many of the following questions are about your parents. If your *own* mother is not bringing you up, answer for your *near-mother*, that is, the woman or girl who is bringing you up or who comes the nearest to it. If your father is not bringing you up, answer for your *near-father*.

68. Does your father (or near-father) work regularly for pay? yes
no
69. How many persons work for either of your parents (or near-parents) full time for pay outside your home? 0
1 or more
70. Have your parents (or near-parents) any money in the bank? yes
no
71. Did your parents (or near-parents) go away on a week's vacation during the last twelve months? yes
no
72. Do you or any of your brothers or sisters who are going to school work after school or on Saturdays for pay? yes
no
73. Have you any money in the bank? yes
no
74. Are you given spending money regularly? yes
no
75. About how often do you play on the sidewalk or street or road? daily
weekly
monthly
never
76. How many children in your family have been taken to court more than once for not behaving? 0
1 or more
77. Do you take paid private lessons in music or dancing or riding or painting or French or anything? yes
no
78. How many of the families living near you have their own yard with trees or bushes or grass? most
half
a few
none

Go on to the next page.

-
79. How many of the men that you know best have jobs? most
half
a few
none
80. How many of the men that you know best (not relatives) have any money in the bank? most
a few
none
81. How many times have you moved in the last three years? 0
1
2 or more

H. YOU AND OTHER CHILDREN

82. How many brothers have you who live with you in your home? 0
1 or 2
3 4 5 6 7 8 9
83. How many sisters have you who live with you in your home? 0
1 or 2
3 4 5 6 7 8 9
84. How many brothers and sisters have you less than 16 years old who are *not* living with you in your home? 0
1
2 or more
85. How many brothers who are living with you in your home are near your own age, not more than two years older or younger than you? 0
1 or 2
3 or more
86. How many sisters who are living with you in your home are near your own age, not more than two years older or younger than you? 0
1 or 2
3 or more
87. How many years have you taken regular care of any young child, one hour or more a day? 0
1
2 or more
88. Do your friends come to see you in your home? often
seldom
never
89. How many hours a week do you play with one or more boys outside your family when you are not at school? 0
2
4 or more
90. How many hours a week do you play with one or more girls outside your family when you are not at school? 0
2
4 or more

Go on to the next page.

91. How often do you play with three or more children at once when you are not at school?
daily
weekly
monthly
never
92. Which child are you most like? *Nan* is with her parents (or near-parents) most of the time. *John* is with children most of the time. *Rose* is with her parents about half of the time and with children about half of the time.
Nan
John
Ben
Rose
93. Which child are you most like? *John* often quarrels. *Mary* seldom quarrels. *Jane* always quarrels.
John
Mary
Jane
Ed
94. Which child are you most like? *Walter* never builds or makes things with other children when he is not at school. *Maud* seldom does. *George* often does.
Walter
Maud
George
Ruth
95. Circle each one that you often talk about with other children when you are by yourselves.
"movies" plants 0
"funnies" animals 1
school children 2
music parents 3
books grown people 4
How many did you circle? 5
6 or more

I. YOU AND YOUR PARENTS

96. Are your parents (or near-parents) well?
both
father only
mother only
neither
97. Do you go to the "movies" with a parent or near-parent?
often
seldom
never
98. How many times do you usually go to the "movies" from Monday morning to Thursday night?
0
1
2
3 or more
99. How many people such as boarders or relatives outside your *own* family live in your home?
0
1
2 or more

Go on to the next page.

-
100. Where did you learn the most about manners? at home
 at school
 somewhere else
101. When you are at home, is your mother (or near-mother) there? usually
 seldom
 never
102. About how many hours a day is your mother (or near-mother) doing things with you when you are not at meals? 0
 $\frac{1}{2}$
 1 or more
103. With what person do you spend the most time? servant
 mother or near-mother
 some other
104. What language does your mother (or near-mother) speak at home most of the time? Japanese
 Spanish
 German
 Russian
 English
 some other
105. Does your mother (or near-mother) give you what you ask for? usually
 seldom
 never
106. Does your mother (or near-mother) praise you for good conduct? often
 seldom
 never
107. How does your mother (or near-mother) treat you? Like a grown person
 little child
 baby
108. How often does your mother (or near-mother) punish you physically? daily
 weekly
 seldom
109. How often does your mother (or near-mother) punish you, not physically but in other ways? daily
 weekly
 seldom
110. Does your mother (or near-mother) let you decide important things for yourself? usually
 seldom
 never

Go on to the next page.

-
111. Does your mother (or near-mother) (a) pay a
little attention to what you do (b) help you b
with the important things you do (c) direct c
everything you do d
112. Does your mother (or near-mother) let you ex- always
plain your conduct before blaming or punishing usually
you? seldom
never
113. Does your mother (or near-mother) take time usually
to answer your questions carefully? seldom
never
114. Circle every one that your mother (or near-
mother) has been doing in the last twelve
months.
Taking a course
Buying books
Reading books
Going to concerts 0
Going to lectures 1
How many did you circle? 2 or more
115. Was your mother (or near-mother) born in the yes
United States? no
116. Does your mother (or near-mother) work regu- yes
larly for pay? no
117. Does your father (or near-father) live in your yes
home? no
118. About how many hours a day does your father
(or near-father) spend doing things with you 0
when you are not at meals? $\frac{1}{2}$ or more
119. When does your father (or near-father) work? day
night
120. What language does your father (or near- Italian
father) speak at home most of the time? Polish
Chinese
English
Yiddish
some other

Go on to the next page.

-
- | | |
|---|--------------------------------------|
| 121. Does your father (or near-father) give you what you ask for? | usually
seldom
never |
| 122. Does your father (or near-father) praise you for good conduct? | often
seldom
never |
| 123. How does your father (or near-father) treat you? Like a
*
grown person
little child
baby | |
| 124. How often does your father (or near-father) punish you physically? | daily
weekly
monthly
never |
| 125. How often does your father (or near-father) punish you, not physically but in other ways? | daily
weekly
monthly
never |
| 126. Does your father (or near-father) let you decide important things for yourself? | usually
seldom
never |
| 127. Does your father (or near-father) (a) pay little attention to what you do (b) help you with the important things you do (c) direct everything you do | a
b
c
d |
| 128. Does your father (or near-father) let you explain your conduct before blaming or punishing you? | always
usually
seldom
never |
| 129. Does your father (or near-father) take time to answer your questions carefully? | usually
seldom
never |
| 130. Was your father (or near-father) born in the United States? | yes
no |
| 131. Does some one person in your family make the important decisions for the others? | always
usually
seldom |

Go on to the next page.

132. Do your parents (or near-parents) make the important decisions in your family by talking things over with the children and deciding together? often
seldom
never
133. Did you go away on a vacation with your parents (or near-parents) during the last twelve months? yes
no
134. How often do you feel that your family likes to have you around? often
seldom
never
135. How often do you feel that you can enjoy things at home without worrying about your faults? often
seldom
never
136. Which child are you most like? *Walter* enjoys meals much more than play. *Nat* enjoys play much more than meals. *Roxy* enjoys both and likes them about equally. *Mary* enjoys neither. Walter
Nat
Roxy
Mary
137. Who is bringing you up? your *own*
mother
guardian
someone else
138. Who took care of you from birth until you began to go to school? your *own*
mother
guardian
someone else

J. YOU AND YOUR COMMUNITY

139. Circle each place where you and your friends meet often and stay an hour or more together when you are not at school.
- swimming pool
theater or "movie"
clubroom
gymnasium
church
- How many did you circle? 0
1
2 or more

Go on to the next page.

-
140. Circle each place where you and your friends meet often and stay an hour or more together when you are not at school.
- | | |
|---------------------------------|-----------|
| street | |
| grocery store | |
| drugstore | |
| cigar store | |
| candy store | 0 |
| other store | 1 |
| <i>How many</i> did you circle? | 2 or more |
141. Have you a library card in your own name at some library, not the school library? yes
no
142. About how many books do you take each month from a library, not the school library? 0
1 or more
143. Circle each one you listen to on the radio almost every day at home.
- | | | |
|---------------------------------|-----------|-----------|
| news | health | 0 |
| comics | science | 1 |
| sports | education | 2 |
| <i>How many</i> did you circle? | | 3 or more |
144. How many of the men that you know best vote when there is an election? most
half
a few
none
145. How many of the women that you know best vote when there is an election? most
half
a few
none

3. THE COMPREHENSIVE ACHIEVEMENT TEST

The *Comprehensive Achievement Test* aims to measure, by sampling, everything important which a child ought to learn and which he can tell in a brief pencil-and-paper test. It is a multiple-choice test used from Grade III through Grade IX¹ and is reliable for classes, grades, schools, and school systems in both subtest scores and total score and, for individuals, in total scores.

An inspection of the content of the test shows subject matter, skills, activities, attitudes, and ideals represented in reasonable proportions and relationships. (Note that some of the titles are used to camouflage the real nature of subtests.)

¹ But it is usable also in high school and college.

lest it cease to be effective and therefore become means to other ends than those sought. It must reveal the strengths and the weaknesses. It must cover the ground of desirable learnings from spelling to character, from technical skill to philosophic insight, from subject matter to activity, from *one plus one* to the art of living, from self-expression to group solidarity and world coöperation. It must be a means of educating the emotions, the wishes, and the intellect. It must do all these things in order that the measurement may no longer cause one-sided emphasis upon whatever happens at the time to be both conventional to teach and feasible to measure. If a test program fails to reveal prejudices or feelings of inferiority, teachers will be less likely to do anything about them. Any program which measures only subject-matter learning and skills tends to keep the school from attempting the transition from any such incomplete education toward education in every aspect of life.

Against such shortcomings, which, after fifteen years of complaint, still characterize programs of measurement, this test is guarded. The influence of every question upon those who read it or give it or answer it—child, teacher, superintendent, school board, publicist—was studied until the test was filled with suggestions for better aims, better methods, better activities. The language of the test suggests concrete ways of becoming better members of society, better friends, better thinkers and appraisers, better leaders, followers, and coöperators, better learners and teachers. Every potentially bad influence that could be found was, if it could be, removed.

Some may think the test is biased either toward progressive education or toward conservative education. Every attempt has been made to avoid prejudice. Purposes constituting the exclusive province of either type were avoided, and common ground was occupied. Everybody, for example, wishes health, reading, self-assurance, coöperation, and enjoyment of life to find their places in the school program. Those educational practices which were approved by one group rather than the other were omitted. If the test leans, the authors do not know to which side. They have tried to make it serve both groups equally, an effort now fortunately facilitated by the two groups themselves, who publicize common goals while using different practices. The express purposes of the two groups have become so much alike that a

single comprehensive achievement test can be made to serve them both.

An unbiased test will be accepted by both sides for the purpose of settling their differences about the practices of education. Representing common elements of purpose, but not practices which differ, the test may help toward a consensus in answer to the questions: "Is it the traditional or the progressive practices which bring us sooner to our common goals? If, for example, the teacher minutely directs the child's work, will the child become more or less responsible? Which will be nearer the common goal after five years, the progressive or the conservative school?"

Some may think that each subtest should measure one, and only one, pure trait. Nothing is known about the amounts of overlapping among subtests in any of the tests of this battery. An arithmetic test will, however, contain the two problems $131\overline{)451}$ and 33×13 , both of which contain the subproblem 3×3 . The two problems, one labeled division and the other multiplication, overlap. A test of foresight and a test of coöperation are bound to overlap, both having an element of intelligence. Nevertheless, it may be foresight and coöperation which we wish to measure. For most purposes we have to measure things which have a part in common. The independent factors thus far isolated by the science of education—ebullience, ability to handle geometric forms, and the others—do not represent the best units of purpose for education to use. They are, therefore, not the best units for most educational tests, special purposes and pure research being excepted. The best units for education and its tests—coöperation, planning, criticism, and others—overlap one another and the statistically isolated independent factors. Since we are primarily interested in the units themselves and not in their independence and since human nature is organized and nurtured as it is, we accept the overlapping. Science is not yet able to build tests made of non-overlapping subtests representing the chief human purposes. It may never be able to do so. It may never wish to do so. If the effects of a test upon human beings are severally good and collectively balanced, we shall be happy.

Some may think that "sophisticated" youth, intent upon high scores, honors, and attendant privilege, not upon abun-

-
4. How often should little children decide what medicine to take? always
usually
often
seldom or
never
5. Which is correct?
(a) $2 + 3 = 1$ (b) $2 + 3 = 5$
(c) $2 + 3 = 2$ (d) $2 + 3 = 3$
a
b
c
d
6. Which child are you most like? John
John looks both ways before crossing the street. Ned
Ned crosses the street without looking. Bill
7. Do you know how to mark the correct answers? yes
no
not sure

Look at me when you finish (*pause*). What is the answer to Question 3? Did you circle two words and also the number 2? What is your answer to Question 4? Why? To 5? To 6? To 7?

On the following pages circle the best answer to *every* question quickly, and without any skipping. When you are told to do so, turn this page and begin working.

TO THE EXAMINER: Read aloud the instructions to the children while they read silently. Make sure that all children understand the meanings of the words *always*, *usually*, *often*, *seldom*, and *never* in Item 4, for they appear frequently in the test. Exactly forty minutes after the children turn the page and begin, collect the test papers and let the children do something else for one-half hour or more. Then redistribute the papers and let the children continue for another and final forty minutes. Give no more help than that already indicated, except to see that the children understand what to do and that they answer every question somehow in so far as possible in the two forty-minute periods.

A. HEALTH AND PLAY

1. How many persons should use the same towel? 1
2
3
2. How can children make a child *like* to be fair in games? a
b
(a) By scolding (b) They cannot (c) By being c
fair to him and to one another (d) By punishing d

Go on to the next page.

-
3. If you see blood spurting very fast from a child's arm, what should you do first?
 (a) Go for help (b) Knot a strip above the wound
 (c) Wrap the arm, covering the wound (d) Call in the police
4. Which child are you most like?
Ned often plays out of doors with other children.
John seldom does. *Alice* plays by herself. *Sue* does not play, but works hard all the time.
5. When people are sick, should they ask the drug-gist what medicine to take?
6. Circle each word naming something that often carries disease germs: fly, rust, oven, water, dust, rat, mosquito, boiling water. *How many* did you circle?
- a
b
c
d
- Ned
John
Alice
Sue
- always
usually
seldom or
never
- none
1 or 2
3 or 4
5 or 6
7 or 8

B. READING

7. When you read to yourself, should you say the words with your lips?
8. Do children grow?
9. In a large city there are a great many
10. Is it usually profitable to sell for less than the cost?
- always
often
sometimes
seldom
- yes
no
perhaps
- rivers
stores
lakes
parks
- yes
no
not sure

11. *Learning to Read*

Some little children tried to read about how to raise chickens. But all the children lived in the city, and none of them had even so much as seen a live hen. They soon discovered that the reading was too difficult for them. Their teacher purchased a brooding

Go on to the next page.

hen and set her upon thirteen eggs. When the eggs were hatched, the children were soon compelled to admit that none among them knew enough to provide a correct diet for the fledglings. So again they essayed the printed information which once had defied them, this time to find that the hard words were as easily interpretable as if a magician had waved his wand over them.

Did the children live on a farm?

ves

no

not sure

12. In the story *Learning to Read* (Question 11) what is the main thing that the writer wanted to tell?
- (a) How to raise chickens
(b) How to learn to read
(c) A story (d) A joke
13. In the story *Learning to Read* how did the children learn to read?
- (a) By reading more (b) By studying the next lesson
(c) By reviewing (d) By reading something they needed to use
14. Which is the best outline of the story *Learning to Read*?
- (a) Reading City Magic
(b) *Learning to Read* The difficulty Experiment What they did
(c) *Reading Made Easy*
Reading hard
Learning by using the reading
Reading easy
(d) None of them

C. FINDING INFORMATION

- 15.** To find *books* on games, use the

Readers'

Guide

publisher

card

catalogue

bookseller

Go on to the next page.

16. To find out what problems your city or town thinks important, go to
 neighbors
 teachers
 neighbors and
 leaders
 library and
 teachers
17. Are these words arranged in the same order as they are in a dictionary?
 flask
 flare
 flaunt
 yes
 no
 not sure
18. In Figure 1 how many years are there between one line and the next?
 0
 5
 10
 15

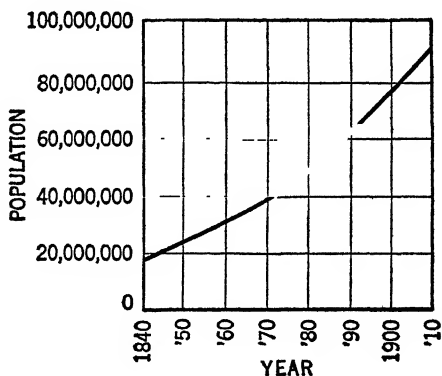


FIGURE 1

19. In Figure 2 about how many millions are there in Canada?
 10
 20
 40
 60
20. In the following table of prices how many months are there between the lowest and the highest prices of sugar?
 1
 2
 3
 4
 5
 6
- | | J. | F. | M. | A. | M. |
|--------|----|----------------|----|----------------|----|
| Sugar | 6 | $5\frac{1}{2}$ | 6 | $6\frac{1}{2}$ | 7 |
| Apples | 9 | 7 | 5 | 4 | 5 |

Go on to the next page.

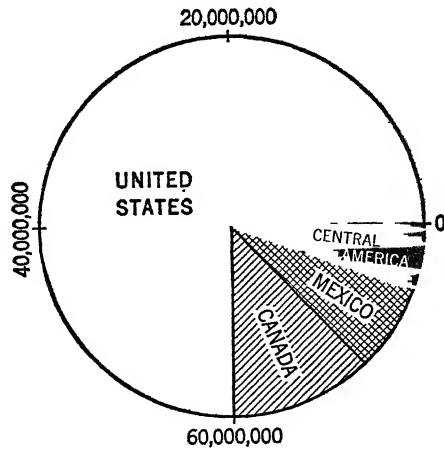


FIGURE 2

D. SPEAKING, WRITING, AND SPELLING

21. Is this sentence correct?
He laid down. yes
not sure
no
22. A good paragraph has how many main ideas? 1
2 or 3
any number
23. In which of these ways will you learn the most important things? (a) Making up a play together as you act it again and again (b) Writing a play and playing it (c) Learning a play and playing it (d) Making money by giving a play a
b
c
d
24. How many sentences do the following words make?
A canoe and three new paddles. 0
1
2
3
25. Circle each word that is spelled correctly.
wheather colonel none
earning occurrence 1 or 2
decided proficient 3
separate 4
How many words did you circle? 5
6
7

Go on to the next page.

E. ARITHMETIC

26. If you buy soap for 7 cents and apples for 5 cents, how many cents must you pay? 2
7
12
35
27. How many minutes from 8:40 to 9:30? 10
40
50
70
28. You had \$9 to spend. You have spent \$1. The rest is for bats at 25 cents each. How many bats can you buy? 2
8
20
32
29. Circle each one that must be on a bank check.
 date none
 name of bank 1
 address of maker 2
 amount in figures 3
 amount in writing 4
 signature of maker 5
 How many did you circle? 6
30. There are fifty-eight teachers in a school and three hundred forty-eight children in the sixth grade. Two hundred thirty-six of these children and teachers are planning to dine together. There are thirteen tables of the same size. How many people should be seated at most of the tables? 4
 $4\frac{6}{13}$
6
 $18\frac{2}{13}$
 $26\frac{10}{13}$
45
71

F. ARTS AND CRAFTS

31. Whittle with a knife (a) toward the body a
 (b) away from the body (c) either way b
 c
32. Which is the best line of poetry?
 (a) And through the moss the ivies creep a
 (b) My spirit longs to flee away b
 (c) I have given you streams to fish in c
 (d) And knit her stockings there d

Go on to the next page.

-

G. UNDERSTANDING THE WORLD IN WHICH YOU LIVE

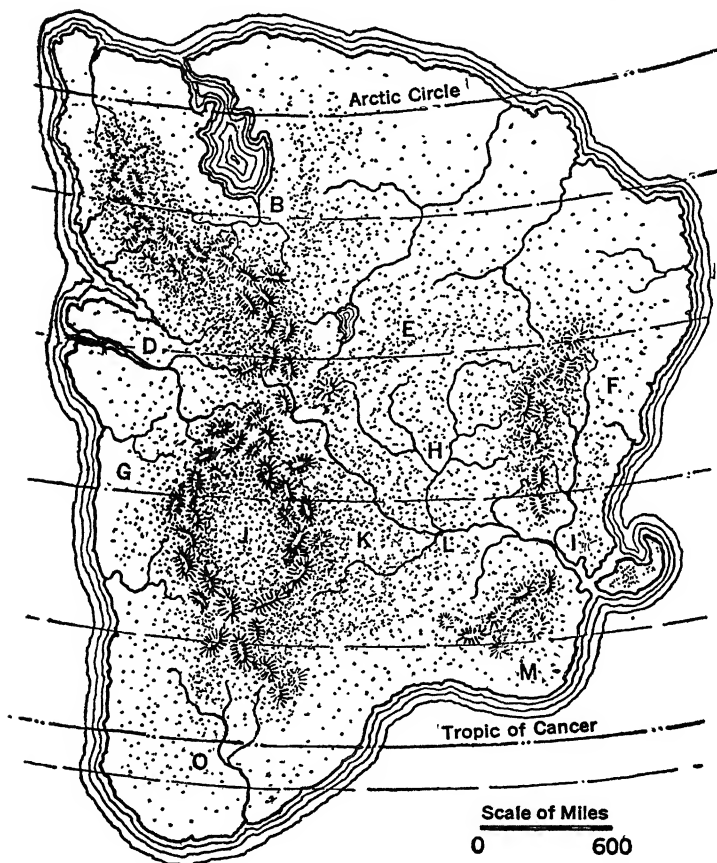
- Go on to the next page.*

41. Destroying forests makes

dust storms
floods
rain
snow

42. Look at the map. Which people find it hardest to work together with other peoples?

I
J
L
M



43. Look at the same map again. Where will the most sheep be carried by rail? From

M to J
B to K
H to F
K to I

Go on to the next page.

44. Most manufacturers try to make goods that will
(a) last forever (b) be used up (c) make the least profit
45. Which is *now* most important to study in history?
(a) The landing of the Pilgrims
(b) Our Civil War (c) How wealth is shared

a
b
c
d

a
b
c
d

H. BUYING AND USING THINGS

46. Is it wise to judge quality by price?
47. Is it wise to buy on the installment plan?
48. How many kinds of fruits and vegetables sold in stores are, while growing, poisoned enough to hurt people who eat them?
49. Advertised cures for colds have often been
50. How many of the following serve consumers on a large scale?
- Consumers League
Consumers Research
Household Bureau
The Coöperatives

always
usually
sometimes
never

always
usually
seldom or
never

none
few
many
all

cheap
sure
dangerous
useless

none
1
2
3
4

I. BEING A SENSIBLE AND A USEFUL CITIZEN

51. How many clubs or societies for games, study, hobbies, or *anything else* do you belong to—outside of school?

0
1 to 4
5 or more

Go on to the next page.

-
52. Which will best show that you are a good citizen? a
(a) Helping to keep the streets clean (b) Reading b
stories (c) Going to see motion pictures c
d
53. Should you have some opinions that you will always
stand by and not consider changing? often
seldom
never
54. Which do you believe?
(a) A man can be loyal to his own country and to
the world at the same time (b) A man's own a
country can do no wrong (c) Your country can b
do no wrong (d) A man should be loyal only to c
his own country d
55. For which one should people be punished the
most?
(a) Making children work in a coal mine (b) a
Making speeches against the government b
(c) Making speeches against socialism (d) c
Stealing ten dollars d
56. What can a child do to help prevent wars? a
(a) Nothing (b) Wait until he grows up (c) b
Get his friends to talk about how to prevent wars c
(d) Just study the problem d
57. People should (a) give unquestioning loyalty to a
public officials (b) work openly against bad ones b
(c) study such problems (d) leave such matters c
to organizations d

J. WATCHING THE PROGRESS OF THE WORLD

58. What do doctors try most to do? a
(a) To keep all the people well b
(b) To cure the sick c
d
59. Will there be any more wars? yes
no

Go on to the next page.

60. Must an increase in production throw men out of work? yes
perhaps
no
61. Which is true?
(a) The unsolved problem is to make enough goods a
(b) The unsolved problem is to get men to turn to the fair sharing of goods and to the art of living b
(c) Both are true c
(d) Neither is true d
62. *An economy of abundance* means (a) produce more a
(b) save goods b
(c) save goods for depressions c
(d) make enough for all d

K. CHOOSING THE BEST EXPERIENCES

63. Which is the best reward in school?
 (a) Money for work (b) Praise for work
 (c) Enjoyment of work
64. Which should you choose?
 (a) \$5 a day friends work health
 (b) \$10 a day friends work poor health
 (c) \$10 a day friends no good work health
 (d) \$10 a day no good friends work health
65. How much freedom should you seek?
 (a) What grown people give you because freedom makes people hard to manage
 (b) Very little
 (c) As much as you can use for everybody's happiness
 (d) Complete freedom
66. Which plan is best?
 (a) Take what experiences come your way
 (b) Plan together how to get the best experiences
 (c) Find out what experiences home and school have planned for you
 (d) Seek every kind of experience

Go on to the next page.

67. Which child said the best thing about the school?

Nan: "That school is not like ours."

Bob: "No. Those children know how to decide among themselves what to do."

Sue: "But they were busy, and the teacher did not have to keep after them." Nan

Ed: "I wonder what the teacher was doing; I did not notice her much." Bob

Sue
Ed

L. TALKING THINGS OVER, HANDLING
DISAGREEMENTS, AND GETTING
THINGS DONE

68. How should the chairs be placed when just a few persons are talking things over?

(a) In a circle (b) In one row (c) In several rows

a
b
c

69. Circle each thing that a good discussion leader does.

Praises good points	none
Encourages timid children	1
Keeps everyone to the point	2
Gives everyone a chance	3
Makes good speeches	4
Comments on each thing said	5
Keeps steering toward something to be done	6
Shows how suggestions can be used	7
<i>How many</i> did you circle?	8

70. How can you best study your quarrels and disagreements?

(a) Think them over (b) Get opinions and try different ways (c) Study them as a lesson (d) Watch people when they quarrel

a
b
c
d

71. What could you do to make a child stop *wanting* to boss other children?

(a) Ask him not to (b) Explain why (c) Talk with him about how it feels to be bossed (d) Boss him until he learns how it feels, and let everybody else refuse to be bossed

a
b
c
d

Go on to the next page.

M. FORESEEING CONSEQUENCES

72. What will happen if you strike another child?
He will (a) cry (b) tell his teacher (c) strike
you (d) feel hurt
a
b
c
d
73. Punishing children who cheat will make them
(a) stop cheating (b) hide their cheating
(c) cheat more (d) cheat less
a
b
c
d
74. What would make children dislike the Chinese?
(a) Hearing grown people speak unkindly of them
(b) The real character of the Chinese (c) Play-
ing with Chinese children (d) Visiting Chinese
homes
a
b
c
d
75. When a boy calls somebody's father an old fool,
what would make him stop?
(a) Fighting him (b) Telling his father (c) Tell-
ing him, "So's your old man!" (d) Letting no-
body notice it
a
b
c
d
76. If children accept the teacher as a member of
their own group, there will be (a) less respect for
the teacher (b) more respect for the teacher
(c) more disorder (d) less learning
a
b
c
d

N. UNDERSTANDING PEOPLE AND THINGS

77. How many white people are brighter than any
other people?
all
some
none
78. How often does your best friend tell the truth?
always
sometimes
never
79. Quickly circle each word you do not like—that
troubles or disturbs or annoys you more than
just a little.
- | | | |
|---------------------------------|---------------|---------|
| Democrat | whisky | |
| cigarette | Republican | |
| Communist | Sunday School | |
| Christian | labor union | 0 to 2 |
| alcohol | capitalist | 3 to 5 |
| Socialist | church | 6 to 8 |
| <i>How many did you circle?</i> | | 9 to 12 |

Go on to the next page.

-
80. Think of the race or nation that you dislike the most. How many persons in it are bad? all
some
none
81. Think of the race or nation that cheats the most. How many persons in it will cheat? all
some
none

O. REMEMBERING THINGS

82. Have you ever before heard of *all* these stories?
Rip Van Winkle
Hans and Gretel
The Three Bears
Little Red Ridinghood yes
The Spellbound Ghost no
83. Read these numbers once. Then say them backward without looking at them.
 4 2 8 7 5 9 1 0 6 4 3 7 yes
 Did you say all the numbers in the right order? no
84. Did you ever take anything that belonged to anyone else, even a pin or a button? yes
no
85. Did you ever act greedily by taking more than your share of anything? yes
no
86. Did you know that a horse named Peter Pan won more than five races? yes
no

P. KEEPING YOUR TEMPER

87. When you have been angry for days and it is hard for you to get over your anger, what should you do?
 (a) Talk the way you feel (b) Play hard, work hard, eat little, sleep (c) Keep from showing your anger (d) Use will power a
b
c
d
88. Your smiling will (a) make you less angry (b) have no effect (c) make you feel angrier (d) stop your anger at once a
b
c
d

Go on to the next page.

89. To learn to get over fits of anger, (a) take some way that your teacher or mother tells you and stick to it (b) try different ways and use what works (c) combine advice with trial (d) think and read about the problem, talk it over, and try different ways

a
b
c
d

Q. MANNERS

90. Open a letter addressed to any other member of your family
91. What are manners for?
(a) To show respect to all (b) To show others that you know how to act (c) To make it easier for grown people (d) To help you to get on with the most powerful people
92. Wait to be recognized instead of interrupting
93. Should you reach in front of another person for food in order to save him trouble?

always
usually
sometimes
seldom

a
b
c
d

usually
often
seldom
never

often
sometimes
seldom
never

R. MODESTY

94. How often do you feel that you can succeed at most things if you have time to learn?
95. How often do you feel that children say many unpleasant things about you behind your back?
96. How often do you feel that children go out of their way to be with you?
97. How often do you feel that both boys and girls like to have you with them?

often
seldom
never

always
often
seldom

often
seldom
never

often
seldom
never

Go on to the next page.

98. How often do you feel that you can enjoy things without worrying about your faults? often
seldom
never

S. ENJOYING LIFE

99. Which child are you most like? Ed
Ed has no friends at all. *Bert* has few friends. Bert
Rose has many friends. Rose
100. Which child are you most like? Ruth
Ruth enjoys her Saturdays much more than her Mondays. *May* enjoys her Mondays much more than her Saturdays. *John* enjoys both days and likes them about equally. *Ned* likes neither day. May
John
Ned
101. If you could do as you like, how many hours a day would you go to school? 2 or less
3
4 or more
102. Which child are you most like? Will
Will enjoys nature study and science much more than social studies, history, and geography. *Mary* enjoys social studies, history, and geography much more than nature study and science. *Bob* enjoys both and likes them about equally. *Rose* enjoys neither. Mary
Bob
Rose
103. Which child are you most like? Dan
Dan enjoys his school much more than his home. *Mary* enjoys her home much more than her school. *Will* enjoys both and likes them about equally. *Jane* enjoys neither. Mary
Will
Jane
104. Which child are you most like? Bill
Bill enjoys his playtime much more than other times. *Ed* enjoys other times much more than his playtime. *Roxy* enjoys both and likes them about equally. *Susan* enjoys neither. Ed
Roxy
Susan
105. What kind of year are you having? a
(a) Sad and gloomy (b) Neither very sad nor very glad (c) Somewhat happy (d) Happy b
c
d

If you finish before time is called, go back over your work and make sure that you have made no mistakes.

4. THE SCHOOL PRACTICES QUESTIONNAIRE

The *School Practices Questionnaire*, Form I of which is given in Chapter XVIII, is designed to measure the extent to which a school has the characteristics of democratic activity. It is used from Grade IV through Grade IX.¹ What is meant in detail by the term *democratic activity* is set forth in the test itself in its many subheadings and questions. Anyone who is fully participating in democratic activity, whether student, teacher, superintendent, member of a school board, patron, or even indirect taxpayer, ought to become more and more competent in the phases of life suggested by the following captions:

SUBTEST TITLES

Facing Situations.....	FS
Living in the Community.....	LC
Discussing Situations.....	DS
Freeing Speech and Thought.....	FST
Freeing Activity.....	FA
Dealing with Conflicts.....	DC
Initiating Activity.....	IA
Planning Activity.....	PL
Evaluating Activity.....	EV
Using Coöperation.....	CO
Motivation.....	M
Using Committees.....	CM
Using Experts.....	XP
Using Books.....	B
Using Knowledge and Skills.....	KS
Using Tools and Materials.....	TM
Using Art.....	A
Using Tests and Experiments.....	TX
Using Records.....	RC
Living Democratically.....	D
Living Happily.....	H

The questionnaire is a test of the curriculum, an instrument with which to evaluate the experiences which the school makes actual for the child. It does not cover all the experiences, but it samples the school's contribution. The titles of the subtests and the questions themselves represent ways in which a school ought to help a child. They comprise and imply the elements of a consistent school program of a certain type. Not everything which a

¹ But it can well be used also in high school and college.

school ought to do for a child can be included in the test, but an essence of such things is identified and illustrated. If we comprise, within its purview, the intended consequences in the hands of sympathetic teachers, as well as the extensions of content in additional forms, the questionnaire richly represents a good curriculum, assuming that a curriculum of democratic activity is good.

The democratic-activist curriculum subordinates knowledge, skills, tests, institutions, liberty, wealth, honor, prestige, and privilege to the fairly-shared good of all. It is activity voluntarily performed by persons and groups for the benefit of each person.

The questionnaire measures the thoroughness with which a school grasps and utilizes the implications of democracy for education. It distinguishes between a democratic school and all others. A school with a higher score is more effectively supporting our indigenous American philosophy of life, while one with a lower score either is less efficient or is working to some other end.

The questionnaire may be used, however, without assuming the value of democracy. A higher score then means that a school is more democratic, without implying that "democratic" is good. If "democratic" is bad, then the lower is the better score.

Some may think that this *yes-no* test ought to have an equal number of *yeses* and *noes* as right answers. It has about four *yeses* to one *no*, but the *noes* often get the vote, two or three to one throughout the test. This makes a larger number of correct *noes* unnecessary and, in fact, undesirable.

As in other tests, the subtests overlap one another; some of the questions can be willfully misused by students; and the reading difficulty necessarily remains too great for a few fourth-grade children.

The *School Practices Questionnaire* is more direct and dynamic than most educational tests. It is more likely than the others to cause growth. It suggests most fully what to do. It combines stimulation, direction, and measurement.

One use of the test is to identify two types of education. The higher scores mean that the school is democratic in type, the lower scores that it is not democratic in type; middling scores suggest mixed types. A wide range of scores in a single class sug-

gests that some of the students have a curriculum of the activity type, while others in the same class do not—a condition which is likely during transition from a non-activity type. A narrow range with a low average indicates a consistently conservative type of school practices, while a narrow range with a high average indicates a consistently democratic type.

Scattered through the test there are questions to which the answers are *no* from the point of view of the activity program. These items represent practices which usually prevail where education is undemocratic—respect for authority rather than for personality, uniform lessons for all rather than individualized social activities, dictated rather than freed activity, and insubordinated subject matter.

The reasons vary which people give for wishing to identify the type of educational practices. The community may wish to satisfy itself that it is getting the kind of education which it wishes, or a research group may wish to make sure that the two types of education it is studying are not mixed but pure.

A *second* use is to measure the excellence of educational practices, that is, the excellence of the curriculum. First it must be decided what type of education shall be considered good. Then it must be determined how closely the curriculum of the school in question approximates that type. The procedure is then like that for identifying the type of educational practices, except that instead of saying that the school is an instance of pure democracy or pure dictation or mixed in type, we wish to speak quantitatively and to say that the school is so good or that it falls so far short of what we wish, or is completely free from dictation, or has far to go before it becomes satisfactorily democratic, or is well along in the transition toward democracy.

A *third* use is to diagnose school practices. The test reveals without any tabulations the weaknesses of a class, whether unhappiness, undemocratic ways, poor work by committees, or want of planning or of coöperation. Pile the papers in order of size of total score, the smallest on top, and then thumb over from paper to paper the subtest scores at the tops of the front pages. Strengths and weaknesses for the class thus appear. The second step is to find individuals who are weakest but who can become stronger. The third step, following diagnosis, is to institute remedial measures, as, for example, to teach students to

solve socially some personal problem by means of a committee for discussion, planning, and work. The fourth step is to check the diagnosis by observing whether the remedy works. If it does, the diagnosis was probably correct. The science of education has not yet advanced enough to check such diagnoses except by means of the success or failure of remedies. Diagnosis must remain uncertain until the remedy is made.

Diagnosis should not be attempted without bearing in mind the limitation inherent in the four-week period to which every question in the test must refer. Check off those questions throughout the test which you think it fair to expect students in general to answer with the better answers for a four-week period. A low score on a subtest is, then, one which is below your estimate thus made; a high score, at or above your estimate.

A *fourth* use is to see how closely the teacher, the principal, etc., agree with the students as to what is going on in the school. After the students have taken the test, let the teacher answer the questions as if she were an average student. If both students and teacher answer truly, they will agree about the strengths and weaknesses as shown by the subtests and by the total scores. If they do not agree, perhaps the teacher thinks she is doing better than she is, or perhaps the students or teacher or both wish to favor the administration in their answers. Whether agreement or disagreement, it should be interpreted. Agreement is to be sought. The testimony of children who are free to answer truly and who try to do so is probably worth as much as that of an expert observer.

A *fifth* use is to measure students' beliefs about what school practices are desirable. For this purpose, do not follow the procedure on the front page of the test paper, but ask the examinees to imagine the best possible school, each from his own point of view. Tell them to change *Did you go to school?* to *In the best possible school, should you go to school?* and similarly for other questions. A higher score means more, and more consistent, belief in democratic activity, a middling score suggests confused belief, and a lower score indicates less belief. Higher scores in the autocratic or dictation items mean less belief in dictation, and lower scores, more. By comparing the dictation score with the scores in the rest of the test, a measure of consistency of belief is obtained. The testing of belief is harder than the testing of

school practices and is more successfully done in the higher grades.

A *sixth* use is to measure the beliefs of teachers, supervisors, principals, and superintendents. For this purpose the test should be taken by such educators in the manner described in the foregoing paragraph. If all take the test, then there is a measure of the agreement or disagreement among the beliefs about the curriculum throughout the school.

An indication of how far the administration has been successful in putting into practice what it believes may be sought by comparing beliefs with practices.

A *seventh* use is to study educational opinions of the board of education. As determiners of policy, they ought to study their own opinions and to lay them before the community. Candidates for election to the board would do well to make a record of the kind of education they cherish and later to boast not necessarily about maintaining their beliefs but, possibly, about growing out of them.

An *eighth* use is to measure the beliefs of school patrons. The parent-teacher association will find the test interesting material for discussion at a series of meetings, especially after those present at each meeting themselves take that part of the test which is to be discussed. An association in which open conversation without trouble is customary can bring teachers, students, and parents nearer together in their views about education and can foster coöperative activity between the school and the rest of the community. However, the whole test is far too much to consider at one meeting.

A *ninth* use is to measure the beliefs of students and teachers of education. How do the beliefs of student-teachers compare with their own beliefs after a few years in the field? Do the students of a certain professor use his ideas in the field? How do his views compare with field practices? Is he ahead of his time? Does he believe in dictation or in democracy, or is he confused? How much does he change the views of his students in one term? How much does he change their practices?

A *tenth* use is to diagnose beliefs. If the students on any level of education are having conservative school practices used upon them, are they content or do they protest? Do they feel dissatisfaction but keep silent about it? Does Alice have a consistent

belief in dictation, or is she breaking away from it? Are the students in substantial agreement about the desirability of using committees?

Does the teacher believe in a different education from that in which the students believe? Does she believe that students ought to participate in planning the daily program, while the students believe that they should not?

Is the teacher democratic in her beliefs, but held by a conservative principal, superintendent, and school board?

Such questions and many others may be studied by comparing scores on subtests with one another, for the various persons engaged together in education, as well as by comparing belief scores with practices scores.

An *eleventh* use is to motivate growth. The stimulation of growth is the most important use of tests in education. Growth can be brought about only through motivation. The language of the test ought to open the educational eyes of students and teachers so that some of them will say, "I did not dream that anyone would be permitted to teach that way." Use the questions as criteria for judging the education in use, and as purposes for suggesting what to do. Live in the community, face the situations there found, help students to do the same, and the resulting growth will affect the score the next time the test is given.

Growth in beliefs is brought about chiefly by discussion, by reading, by contacts with novel ways of life, and by trial. The obvious and sound methods of causing growths in beliefs about education are, therefore, such as class discussion about the questions in the test, reading of diverse views on educational problems, visiting schools conducted differently from one's own, and trying new ways of educative living. Better if the four ways shall proceed together. Let students trying for the first time to use committees read about using committees, discuss test questions on using committees, and visit committees in some other school.

Certain dangers await those who experiment with growing beliefs, some of which may possibly be avoided by heeding the following advice. Pressure, whether from fear, undue respect for authority, or desire to please friends, makes for unhealthful changes of views. Sudden growth is not to be desired. People

should be allowed to discuss issues and then go home with their minds unsettled. They should not be pressed for the decision, but it should come as it will after the educative exposures have been made.

A *twelfth* use is to set expectancy standards for school practices. How much progress is it fair to expect a class to make in a month? The teacher can set her own standard, though science cannot do it for her; and the best way is to do it experimentally. First give the test, following the directions on the front page. Then study the answers of the students and count those answers which are wrong but which you think you can change to right for most of the students during the month. Add that number to the class average score. This gives the expectancy standard for the class. This standard, like the diagnosis already discussed, can be verified only by trial. If you can change the scores that much, the standard is probably low enough. Modify the school practices in your room. Abandon dictation, uniformity of lessons, dependence upon textbooks, and too many rules and regulations, and begin meeting needs which children themselves already feel or which you can wisely and soon help them to feel. Begin to teach students to do better than before those desirable things which they are likely to do anyhow. After a few trials you can come to set standards with a more realistic accuracy than before. Note how much gain you thus make, for example, in October of two successive years and how closely you can estimate how much to undertake in each four-week period. The setting of such standards is thus intimately and necessarily connected with the class itself and cannot be done apart from it.

To set standards for the growth of beliefs is a harder task, and yet it may be attempted in the same manner as for practices. The earlier guesses at the amount of growth that will take place may be poor, but they can be bettered by repeated trial. After a few successive months, not to say years, the teacher will come to know something about how long it usually takes children to change their ideas about what kind of education they wish.

A *thirteenth* use is to measure growth. Growth is measured by comparing scores at intervals. In school practices, growth is sometimes comparatively rapid. A teacher of children whose activity has always been dictated, who is free to act, who knows how, and who wishes to act, can in a few weeks make a large in-

crease in the scores of her students. Another teacher of like children, not free, not knowing how, not wishing to act, may never increase the scores. The best interval of measurement, depending more upon will, freedom, and competence than upon finance, crowding, and students' intelligence, may vary from once a month to once a term. If the children will tell the truth even after they know the right answers (the proper use of the test will motivate fidelity of report), they may use the same form of the test over and over again until it can suggest little further by way of improvement. Give the same form monthly or at longer intervals until no further gain is made because the ways of making gains are exhausted. Use other forms when new stimulation is needed.

If the class score seesaws up and down from month to month, perhaps the interval is too short, some of the answers are false, or effort is impulsive or inconsistent. Those within the situation can judge which one it is. Verify scores, increase intervals, stabilize effort, and the inconsistencies will smooth out. If the answers are true, a regularly ascending curve is evidence of growth.

Standards of progress in this test cannot be transferred from group to group without weighing four factors: truth, wish, freedom, and ability. Again, those within the situation can best judge which of these factors prevent measured growth. Unless the answers are true, the picture is beclouded. Growth cannot be forced upon the unwilling. Those having authority can facilitate or hinder growth as they may choose. Ability to grow is sometimes rapidly achieved. The pupils themselves never preclude good education, except individually and for a time. The determination of quantitative standards of school practices applicable to any class would involve weighing more factors than we can yet measure. Each class must, therefore, make its own standards. If a standard is barely reached on time, it is well set. If it is not reached or is over-reached, it may be better set next time. It is worth more to make all the possible progress and to evidence the amount made than it is to predict the amount. A standard set to be reached at a certain time remains interesting only so long as it does not degenerate into a competitive goal with ends outside the lives of the children, or, once reached, a block to further effort. The most useful standards are just the ones which the test can set: a good school is coöperative, not

competitive or coercive; it is planning, not letting things go at their own gait; it is critically insistent, not unthinkingly acquiescent. Of such standards, the test is full. Amounts of progress ought to be watched and made on time, but motivation of progress ought to be intrinsic; people ought to give their minds first to human perplexities and troubles and incidentally and afterward to growth curves.

5. CONCLUSION

This series of tests measures more completely than heretofore the child's educability, achievement, and curriculum.

The education of the child depends upon the influence of his school, upon the influence of his home and his community, and upon himself. These factors and their consequences constitute the child and his environment. They are exactly that which must be understood and controlled by those who would assure to the child his best possible life. The amount of a child's future educational achievement can be foretold for the next term or year when we know his education hitherto, his capacity for further education, the educative influences of his home and the rest of the non-school community, and the educative influence of his school.

The influence of the school upon the child is limited by the child's intelligence, his past learning, and his non-school environment—factors not usually affected by effort. Intelligence can be injured by disease and perhaps by disuse and ill-advised activities. The past education cannot be changed, though it can be utilized, more or less, for future good. Home and community are not often affected by the quality and impingement of their school, though some have been notably benefited thereby. Taken together, the child's capacity to learn, his present total educational achievement, and his home and community have been relatively fixed assets and liabilities.

Despite these limitations of the school's influence upon the child, and even within them, there is much that the school can do. A child's character does not usually depend upon his being brilliant as a student or upon his being wealthy. The brighter, better-to-do children will use more complex, more effective patterns of response; but the duller, worst-to-do children are usually as ready as any to coöperate, show courage, work hard, be fair,

and live democratically and happily. A child's subject-matter learning does, however, depend upon both intelligence and economic levels, but even here the variations among equally bright and equally wealthy children are large.

How much difference the school makes to the child depends in part upon the type of education it uses. The more the school approximates the activity type, experimental in essence, the fewer are the arbitrary limits to change. Bent upon dictation and the factual grind and measuring for the most part only subject matter and skill, a school is relatively impotent for good, if not actually pernicious. The existence of programs and tests whose components reflect a more complete and better proportioned social life opens before the school entrancing vistas of new and newly measurable achievements.

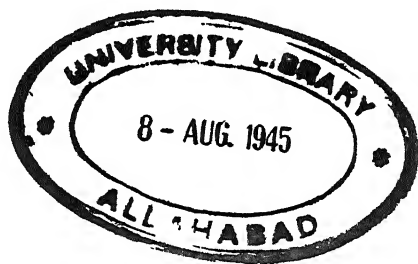
The school has not been able to define for itself its responsibility for the education of the child. Teachers did not know how far it was possible to be responsible because they did not know how much relative allowance to make for non-school influences, past education, and intelligence. Since there was no measure of non-school influences, the burden had to fall disproportionately, wholly, upon the others. Children were not the only ones to suffer from this condition, for teachers took unjust criticism and loss of prestige and income because the imponderability of social and economic differences then stood in the way of assigning to any of the factors its true importance.

But the school can now measure more nearly what it needs to measure in order to complete the requisite knowledge of its situation. It can, by means of the very instruments of measurement, more readily set in motion the influences needed for democratic education. It can more clearly foresee the consequences of its efforts. It can more wisely set standards for improving both its curriculum and the achievements of its students. And it can more discriminatingly appraise its work.

The four tests stressed in this chapter may be used in the elementary, junior high, and senior high schools, and the grade score and age score techniques for using them, described in the next chapter are generally appropriate for all three levels for any tests in the primary, elementary, junior high and senior high schools. The appropriateness of the techniques to the senior high school, especially, assumes that the tests selected for use are

those which measure those fundamental traits with which high schools should be primarily concerned. Unfortunately most high schools tend to be so preoccupied with inculcating and testing mere knowledge of subjects, which may be placed in any year, that the grade score and age score techniques are not so useful. They must usually give place to crude scores and comparison with crude score norms for the number of semesters the subject has been studied. However, the manual accompanying the test in question usually provides adequate directions to guide the user.

The grade score and age score techniques are entirely appropriate to the primary school, but the tests described in this chapter are too difficult, although a plan is suggested which permits the use of the *School Practices Questionnaire* in the primary grades. Similar tests of the right difficulty may be selected from the list given in Chapter VII.



CHAPTER XV
INSTRUCTIONS FOR USING THE
COMPREHENSIVE TESTS¹

1. THE INTELLIGENCE TEST

Administration and Scoring

1. Administer the test, using the directions on the front page.

2. Score the test, using the following directions:

Omit tests, unless completed, of children not present all the testing time, writing the word INCOMPLETE boldly across the front page of each such paper. Open all the remaining papers to Question 1. Copy the correct answers from Column 54 of the Record Sheet upon an unused test paper in the spaces where students write the answers. Check to make sure that there are no mistakes in copying. Make scoring stencils by folding along the right-hand edge of the copied answers. Use these stencils for scoring, marking each item with a dash for right and a zero for wrong or omitted. Mark correct any answer which shows that the child knew which word or number does not belong with the other four. If in doubt, mark it wrong. Count the number right for each child, recording it on the front page. Arrange the papers alphabetically by last names. Have the scoring, counting, recording, and alphabetizing checked by a second teacher. Write his or her name in the upper right-hand corner of the first paper.

Preparation of Record Sheet

3. Take a Comprehensive Record Sheet. Fill the blanks at the top of the sheet. Enter the students' names, alphabetically by last names, in Column 6 of the Record Sheet, without skipping any lines. The number beside the last student's name will then be the number of students.

¹ Quoted or adapted in part from *A Comprehensive Test Program-Manual for Teachers* by William A. McCall and John P. Herring, and with the kind permission of Laidlaw Bros., Chicago.

Determination of Grade Norm

4. Convert each pupil's grade membership into the G score expected of him in view of his grade, i.e., into a G grade. This will be the same for all pupils unless the class contains more than one grade. To do this, use the pupil's grade as the integer and the number of months he has been in the grade as the decimal. Thus 4 high on December 12 is 4.8; 6 low on October 20 is 6.2; and 6 high on April 1 is 6.7. Record each G grade in Column 7 of the Record Sheet. Add Column 7. Do not include a pupil's G grade in the total unless he has a score on *all* tests which were administered. Record the total. Divide by the number of students whose G scores enter into the total. Record the quotient to the nearest tenth. Record, for example, 4.4, not 4.35 or 4.44.

4A. If it is preferred to use age scores¹ throughout rather than G scores, first determine the G grade as in Step 4. Then convert the G grade into its equivalent age, using Columns 5 and 1 of the Record Sheet. Thus a G grade of 5.9 becomes an expected age score for the grade of 11.2. It is the same for all pupils in the class. Record it in Column 7. Compute and record the total and average of Column 7, as in Step 4.

Determination of Age Norms

5. Convert each pupil's age in years and months into an age in years and tenths, using this table:

Months.. . . .	0	1	2	3	4	5	6	7	8	9	10	11
Years.....	.0	.1	.2	.3	.3	.4	.5	.6	.7	.8	.8	.9

Then convert this age in years and tenths into the G score expected of the pupil in view of his age, i.e., into a G age, using Columns 1 and 5. Thus 10 years and 7 months becomes 10.6 years, which in turn becomes a G age of 5.3. Record each G age in Column 8. Compute and record the total and average of Column 8 as in Step 4.

5A. If it is preferred to use age scores throughout rather than G scores, convert each pupil's age in years and months into his age in years and tenths, using the table in Step 5. Record each age, so obtained, in Column 8 of the Record Sheet. Compute and record the total and average of Column 8, as in Step 4.

¹ Steps 4A, 5A, 7A, etc., will usually be omitted.

6. How old or how young for the grade is each pupil? The whole class? Compare the G scores (or age scores) in Columns 7 and 8. If the pupil's G grade is larger than the G age, the pupil is young for his grade, has probably been promoted faster than average, and has probably the capacity for intellectual leadership within his grade. Similarly for the class.

Grade or Age Scores in Intelligence

7. Convert each pupil's intelligence score (number right) into a G score for intelligence, using Columns 2 and 5. Thus an intelligence score of 60 becomes a G intelligence of 4.8. Record the G intelligence in Column 9. Add Column 9. Include in the total only the scores of pupils present for all tests which were administered. Record the total. Divide by the number of pupils whose scores enter into the total. Record the quotient to the nearest tenth. Record, for example, 4.4, not 4.35 or 4.44.

The intelligence G score represents the level of ability to do intellectual work. In that respect it resembles mental age, while it differs from the latter by being a G score rather than an age score. It corresponds to the mental age, not to the intelligence quotient. A G intelligence of 5.0 means the average mental level of children at the beginning of the fifth grade, whether in September or in February. The G intelligence of 5.1 means the mental level of average children one tenth of a grade later. The average at the foot of the sheet is the class G score for intelligence and is interpreted in the same manner as a pupil's G score for intelligence, except for being an average score, not a score of some individual child.

7A. If age scores are used, convert each pupil's intelligence score into a mental age, using Columns 2 and 1. Thus an intelligence score of 60 becomes a mental age of 10.1. Record these mental ages in Column 9. Add Column 9. Compute and record the total and average, as in Step 7. A mental age is the level of intellectual ability which is average for children of that age.

Interpretation of Intelligence Scores

8. How much is each pupil's, or the class's, G score (or age score) for intelligence above or below the grade norm, i.e., what is expected in view of the grade? Compare the scores in Columns 7 and 9.

9. How much is each pupil's, or class's, G score (or age score) for intelligence above or below the age norm, i.e., what you expected in view of the chronological age? Compare Columns 8 and 9.

If age scores are used, this comparison is often expressed as an intelligence quotient (I.Q.). The intelligence quotient is the rate of growth of mental age, the average rate being taken as 100. An I.Q. of 125 means growth in mental age which is 25 per cent faster than the average rate, while an I.Q. of 75 means growth only 75 per cent as fast as the average rate.

If I.Q.'s are desired, divide each pupil's mental age in years and tenths by his chronological age in years and tenths, using no divisor larger than 20.0, carrying the quotient to the nearest hundredth and multiplying by 100. Use no decimal points in writing I.Q.'s. Record the pupil's I.Q. on the Record Sheet at the end of the pupil's name. Compute and record the total and average, as in Step 7.

10. Send to the principal the set of scored papers and, separately, the papers, if any, marked incomplete.

2. THE EDUCATIONAL BACKGROUND QUESTIONNAIRE

Administration and Scoring

11. Administer the questionnaire. If it is felt that some parents may resent the school's asking any of the questions in the questionnaire, identify the test papers in the following manner.

Before giving the papers to the children, but not in the presence of any child, open a paper to Question 1, marking a dash in ink immediately after the question mark. This paper is for the first child. Open a second paper to Question 2, marking that question in the same manner. That paper is for the second child. So mark the papers for the remaining children, numbers 3, 4, 5, etc. On a sheet of paper write in a column at the left the numbers 1, 2, 3, 4, 5, etc., as many as there are children. Plan the exact order in which the papers will be distributed to the children, whether down one row and up the next or in some other order. Opposite each number on the sheet of paper write the name of the child corresponding to it.

Tell the students not to write their names on the test papers and make sure that no student does. (Thus protect parents

against tests being lost or exposed before they reach the principal's office. If, after all, some parent does protest, destroy in his presence his child's test paper.) Distribute the test papers with your own hands in such a manner that each child will have the right one.

Then follow the directions on the front page of the test and let the children begin. While the first two pages are being answered, move quietly from test paper number 1, in numerical order, to the last one, assuring yourself that each child has the right paper. If you find children who have the wrong papers, do not exchange papers, but exchange names on the sheet of paper bearing numbers and names. See that every test is completed.

12. Score the test, using the following directions:

Omit papers not substantially completed, writing the word **INCOMPLETE** boldly across the front page of each. Open all the remaining test papers to Question 1. Make stencils for scoring as for the Intelligence Test (see Step 2), using Column 55 of the Record Sheet. Mark each question with a dash for right or a zero for wrong or omitted. Score the first column for all papers, then the second for all papers, etc. Count the number right for each subtest of the first paper, recording each subtest score in the appropriate blank on the front page. Add the scores of the subtests for each child, recording the total in the appropriate blank. Arrange the test papers in alphabetical order, using the sheet of paper bearing the names. Have the scoring, counting, recording, adding, and alphabetizing checked by a second teacher. Write his or her name at the top of the first paper in the upper right-hand corner.

Grade or Age Scores in Background

13. Take a Comprehensive Record Sheet and fill blanks as in Step 3, unless already done for the Intelligence Test. Convert background scores into G scores for background, using Columns 3 and 5. Record each G score in Column 10 of the Comprehensive Record Sheet. Compute and record the total and the average of Column 10, as in Step 7.

13A. If age scores are used, convert background scores into age scores for background, using Columns 3 and 1. Record the age scores in Column 10. Add, average, and record, as in Step 7.

14. The background G score represents educability not only in intellectual but also in social matters. It correlates with grade status because it correlates with intelligence and because older children do better in social matters than younger ones. It is like mental age because it measures educability and because it increases from grade to grade. It differs from mental age in being a G score, not an age score, and in measuring something much broader than intellect.

Interpretation of Background Scores

15. How much is each pupil's, or each class's, G score (or age score) for background above or below the grade norm? Compare the scores in Columns 7 and 10.

16. Are pupils and classes with better backgrounds also more intelligent? Compare Columns 9 and 10.

17. Are pupils and classes with better backgrounds also brighter? Compare Column 10 with the difference between Columns 8 and 9 or, what amounts to much the same thing, with the I.Q.'s.

18. How does each pupil, or the class, compare with other pupils, or with other classes? See Column 10.

Expectations of Achievement

19. How can the pupil's G score for intelligence (Column 9) be combined with the G score for background (Column 10) so as to determine a reasonable expectation of achievement? Use the formula

$$\frac{2Gi \text{ plus } Gb}{3}$$

If used, this combination expectancy score can be recorded in the margin of the Record Sheet. In order to fix an appropriate expectation of achievement for one or two or three months, etc., later, add respectively 0.1 or 0.2 or 0.3, etc., to the quotient obtained by using the formula. If a class is above average in intelligence quotient, add more; if below, less.

Some Further Suggestions

20. Study the test papers item by item in order to become acquainted with the pupils and to understand their problems.

21. Just before delivering the papers to the principal, write each pupil's name on his paper, but not in the presence of children.

22. Deliver in person to the principal the set of scored papers and, separately, the papers, if any, marked incomplete. They should be filed at once where they cannot be accidentally or curiously observed.

3. THE COMPREHENSIVE ACHIEVEMENT TEST

Administration and Scoring

23. Administer the test, following the directions on the front page.

24. Score the test according to the following directions:

Omit test papers, unless completed, of children not present all the time for both forty-minute periods, writing the word INCOMPLETE boldly across the front page. Open all the papers to subtest A. Memorize the officially correct answers (Column 56) for that subtest. Score that subtest for the first paper, marking each item with a dash for right or a zero for wrong or omitted. Score as right every item in which a pupil indicates in any way that he knew the right answer, but score as wrong every item in which more than one answer is circled. Record the number right on the front page of the test in the appropriate blank. In the same manner score and record subtest A for the rest of the students. Proceed in the same way for the rest of the subtests. Add the scores of the subtests for each student, recording the sum in the space provided on the front page. Arrange the papers alphabetically. Have the scoring, counting, recording, adding, and arranging of papers checked by a second teacher. Write his or her name in the upper right-hand corner of the first paper.

Grade or Age Score in Achievement

25. Take a Comprehensive Record Sheet and fill the blanks as in Step 3, unless already done for the Intelligence Test or the Educational Background Questionnaire. Convert each pupil's achievement score for the whole test into a G score, using Columns 4 and 5 of the Comprehensive Record Sheet. Record the G scores in Column 11. Compute and record the total and the average, as in Step 7.

32. In each subtest how much is the class above or below what is expected in view of its age? Compare the class G score (or age score) in each subtest with the average of Column 8.

33. Send to the principal the set of scored papers and, separately, the papers, if any, marked incomplete.

34. In each subtest how much is the class above or below what is expected in view of its intelligence? Compare the class G score (or age score) in each subtest with the average of Column 9.

35. In each subtest how much is the class above or below what is expected in view of its background? Compare the class G score (or age score) in each subtest with the average of Column 10.

4. THE SCHOOL PRACTICES QUESTIONNAIRE

Administration and Scoring

36. Administer the test, following the directions on the front page. The test is valid only for children who have been in the same class or classes with the same teacher or teachers for four weeks preceding the test. If a child takes the test before those four weeks are up, his paper should be labeled and interpreted accordingly. Newcomers may take the test as soon as they have been present four weeks. Children present less than sixteen school days should wait. This limitation does not apply to other tests or to this one when it is used to measure beliefs.

37. Score the test, following these directions:

Omit papers not substantially complete, writing the word INCOMPLETE boldly across the front page of each. Arrange the remaining papers in alphabetical order by last names. Column 57 gives the items for which *no* is the correct answer. Look at the first such item in the first test paper, Item Number 2. If the pupil circled *no*, circle with a colored pencil the *yes* beside it. If the pupil circled *yes* for this item, mark the *yes* with a heavy colored cross. Do the same for all the items for which *no* is the correct answer, and for no others. For each subtest, count the number right, i.e., all the circled *yeses* which are not marked with colored crosses, no matter whether you or the pupil made the circles. Record, in the blanks on the front page, the total for each subtest. Add the subtest scores and record the total in the space provided for it. Proceed in the same manner for the other

papers. Have the scoring, counting, recording, adding, and alphabetizing checked by a second teacher. Write his or her name in the upper right-hand corner of the front page of the first paper.

Recording and Interpretation of School Practices Scores

38. Take a Comprehensive Record Sheet and fill the blanks as in Step 3, unless previously done for some one of the other tests. Record the total scores in Column 31 of the Comprehensive Record Sheet. Compute and record the total and average of Column 31, as in Step 7.

The interpretation of the practices questionnaire is different from that of the other tests. The school practices in a fourth grade ought, of course, to be just as good as those of any higher grade. Yet the practices of the higher grades can and should be more complex, more difficult. Hence the higher grades can and should attain higher scores. Upper grades permit higher scores than lower ones after all grades have for several years striven equally to be democratic. The overlapping of low with high grades is, however, so large that a fourth grade may average higher than the ninth in the same building. Grade norms have, therefore, less significance than in other tests; the important matter here is not so much comparison with the scores of other classes or schools as with the scores set up by the school as aims. Teachers may from month to month estimate what scores their classes should reach, and they may also study the completed test in order to discover individual weaknesses and strengths and new modes of activity. For other uses, refer to the discussion of the *School Practices Questionnaire* in Chapter XIV, 4.

5. GRADE, SCHOOL, AND SCHOOL-SYSTEM AVERAGES

39. Take a Record Sheet. Record in Column 6 the class designations of all the classes measured in a grade or half-grade in the elementary school, giving one line to each class. Do not record more than one grade or half-grade on the same sheet. Record in Columns 7 through 31, allowing a line for each class, all the G score averages (or age score averages) from the Record Sheet for each class.

40. Add vertically each of Columns 7 through 31. Record the totals and divide each total by the number of classes which enter

into it, as in Step 7. Record the averages. Here and hereafter each class is given equal weight, regardless of the number of pupils in it. This equal weighting, despite variation in number of pupils from class to class, may be avoided, if it is so desired, by recording for each class, totals instead of averages. If this is done, the total of the totals should be divided by the total number of pupils involved, instead of by the number of classes.

41. Interpret these averages as in Steps 6, 8, 9, 15, 16, 17, 18, 19, 26, 27, 28, 30, 31, 32, 34, and 35.

42. Proceed in like manner for each grade or half-grade, using a separate sheet or section of a sheet for each.

43. In order to study a whole elementary or high school, proceed as in Steps 39 through 42, recording average scores for grades or half-grades instead of average scores for classes.

44. In order to study a grade or half-grade throughout a school system, proceed as in Steps 39 through 42, recording average scores for all classes in a given grade or half-grade throughout the school system.

6. RELIABILITY AND NORMS

The table for reading intelligence and achievement G scores for Form 1 of each test was developed upon approximately 20,000 New York City pupils in Grades Four through Nine. The background G scores for Form 1 were developed upon about 6000 of the same 20,000 pupils, also in Grades Four through Nine. The half-grade means were plotted and the curves drawn, smoothed, and extrapolated. Adult norms were secured in order to guide extrapolation.

Data ¹ from two other cities appear in the Record Sheet.

The reliability of a pupil's total score on each test, expressed in indexes of reliability and in probable errors of scores, is given in Table 20. The indexes of reliability tell the correlation between one fallible test and the truth. A coefficient or index of 1.0 indicates perfect correspondence.

But the best and most meaningful indication of reliability is the probable error of a pupil's score by the formula

$$\text{P.E. score} = 0.6745 \text{ S.D. distribution } \sqrt{1 - r_{11}}$$

¹ Omitted in this book.

TABLE 20

INDICES OF RELIABILITY FOR INTELLIGENCE TEST, EDUCATIONAL BACKGROUND QUESTIONNAIRE, AND COMPREHENSIVE ACHIEVEMENT TEST

	$\sqrt{r_{II}}$ INDEX OF RELIABILITY IN AN AGE GROUP	P. E. OF A G SCORE		
		Grade 4	Grade 6	Grade 8
Intelligence.....	.97	.4	.6	.6
Background.....	.91	.8	1.0	.7
Achievement.....	.96	.8	.4	.6

Thus the fourth-grade P.E. score of .4 for the Intelligence Test means that a pupil's obtained score probably differs from his true G score by 0.4 G, once in two times, but practically never differs from his true score 1.8 G. Hence the pupil's score may be assumed to be roughly accurate within 1.0 G. The *Stanford Revision of the Binet-Simon Scale* has about the same reliability.

7. CLASSIFICATION, PROMOTION, AND SECTIONING

The technique for classifying pupils to whom these tests have been administered is the same as that described in Book Three and need not be repeated here. Until more is known about how to weight each test, the following formula may be used for computing G_p:

$$G_p = \frac{2 G_i + 1 G_b + 4 G_e}{7}$$

If desired, G_t may be added to the formula with such weight as is advisable.

TABLE 21

GRADE SCORES, AGE SCORES, NORMS, LISTS OF ANSWERS AND RECORD SHEET FOR A COMPREHENSIVE TEST PROGRAM

COMPREHENSIVE RECORD SHEET														
1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
AGE	INTELLIGENCE	BACK-GROUND	ACHIEVEMENT	G SCORE	AGE	INTELLIGENCE	BACK-GROUND	ACHIEVEMENT	G SCORE	AGE	INTELLIGENCE	BACK-GROUND	ACHIEVEMENT	G SCORE
5.0		0-47	0-4	0.0	10.8	68	76	44	5.5	15.8	106	107		11.3
5.1		48	5	0.1	10.9	69		45	5.6	15.9			70	11.4
5.2	0		6	0.2	11.0	70	77		5.7	16.0	107			11.5
5.3	1	49	7	0.3	11.1	71		46	5.8	16.1		108		11.6
5.4	2-3		8	0.4	11.2	72	78		5.9	16.2		109	71	11.8
5.5	4	50	9	0.5	11.3	73		47	6.0	16.3	108			11.9
5.6	5-6			0.6	11.4	74	79		6.1	16.4		110		12.0
5.7	7	51	10	0.7	11.5	75		48	6.2	16.5	109		72	12.1
5.8	8		11	0.8	11.6		80		6.3	16.6		111		12.2
5.9	9	52	12	0.9	11.7	76		49	6.4	16.7		112		12.3
6.0	10-11			1.0	11.8	77	81		6.5	16.7			73	12.4
6.1	12	53	13	1.1	11.9	78		50	6.6	16.8	110	113		12.5
6.2	13-14		14	1.2	12.0		82		6.7	16.9				12.6
6.3	15	54	15	1.3	12.1	79			6.8	16.9		114	74	12.7
6.4	16-17		16	1.4	12.1	80	83	51	6.9	17.0		115		12.9
6.5	18	55	17	1.5	12.2	81			7.0	17.1	111		75	13.0
6.6	19		18	1.6	12.3		84	52	7.1	17.2		116		13.1
6.7	20	56		1.7	12.4	82	85		7.2	17.3		117		13.3
6.8	21-22		19	1.8	12.5	83		53	7.3	17.4	112		76	13.4
6.9	23	57	20	1.9	12.6	84	86		7.4	17.5		118		13.5
7.1	24		21	2.0	12.7			54	7.5	17.6		119	77	13.7
7.2	25-26	58	21-22	2.1	12.8	85	87		7.6	17.7	113	120		13.9
7.3	27		23	2.2	12.9	86			7.7	17.8			78	14.0

7.4	28	59	24	2.3	13.0	87	88	55	7.8	17.9		121	79	14.1
7.5	29-30			2.4	13.0				7.9	18.0		122		14.3
7.6	31	60	25	2.5	13.1	88	89	56	8.0	18.1	114			14.4
7.7	32	61	26	2.6	13.2	89	90	57	8.2	18.2		123	80	14.5
7.8	33-34	62	27	2.7	13.3	90	91	58	8.3	18.3	115	124		14.7
7.9	35		28	2.8	13.4	91			8.4	18.4		125	81	14.9
8.0	36			2.9	13.5				8.5	18.5				15.0
8.2	37-38	63	29	3.0	13.6	92	93	59	8.6	18.6		126		15.1
8.3	39	64	30	3.1	13.7	93	94	60	8.8	18.7	116	127	82	15.3
8.4	40		31	3.2	13.8	94			8.9	18.8				15.4
8.5	41-42	65		3.3	14.0	95	96		9.0	18.9		128	83	15.5
8.6	43			3.4	14.1		97	61	9.1	19.0				15.6
8.7	44	66	32	3.5	14.2	96			9.2	19.0	117	129		15.7
8.8	45-46		33	3.6	14.3	97	98	62	9.3	19.1		130	84	15.9
8.9	47	67	34	3.7	14.4	98			9.4	19.2		131		16.0
9.0	48-49		35	3.8	14.5				9.5	19.3	118	132	85	16.2
9.1	50			3.9	14.5				9.6	19.4		133		16.4
9.3	51	68	36	4.0	14.6	99	99	63	9.7	19.5		134		16.6
9.4	52	69	37	4.1	14.7	100	100		9.8	19.7	119	135	86	16.8
9.5	53-54		38	4.2	14.7				9.9	19.8				16.9
9.6	55	70		4.3	14.8	101	101	64	10.0	19.9		136	87	17.0
9.7	56			4.4	14.9	102	102		10.1	20.0	120	137		17.2
9.8	57	71	39	4.5	15.0	103	103	65	10.2	20.1		138		17.3
9.9	58	72	40	4.6	15.1				10.3	20.1		139	88	17.4
10.0	59		41	4.7	15.2	104	104	66	10.4	20.2	121			17.5
10.1	60	73	42	4.8	15.3	105	105		10.5	20.3		140		17.6
10.2	61-62	74	43	4.9	15.3		106	67	10.6	20.4				17.7
10.3	63	75		5.0	15.4	106			10.7	20.4	122	141	89	17.9
10.4	64			5.1	15.5			68	10.8	20.5		142		18.0
10.5	65			5.2	15.6				10.9	20.6		143	90	18.1
10.6	66			5.3	15.6			69	11.0	20.7				18.3
10.7	67			5.4	15.7				11.2	20.7				18.4

TABLE 21—Continued

COMPREHENSIVE RECORD SHEET										
City or Town			Grade		School					
1	2	3	4	5	6	7	8	9	10	11
AGE	INTELLIGENCE	BACK-GROUND	ACHIEVEMENT	G SCORE	NAMES OF PUPILS (or classes or grades or schools)	DATES OF TESTS		INTELLIGENCE	BACK-GROUND	ACHIEVEMENT
						GRADE	AGE			
20.8		144		18.5	1					
20.8		145		18.6	2					
20.9			91	18.8	3					
21.0	123			18.9	4					
21.2			92	19.3	5					
					6					
21.3	124			19.4	7					
21.4				19.6	8					
21.4			93	19.7	9					
21.5	125			19.8	10					
21.7	126		94	20.2	11					
					12					
21.9	127			20.7	13					
22.2			95	21.2	14					
22.3	128		96	21.3	15					
22.5			97	21.7	16					
22.6	129			21.8	17					
					18					
22.8			98	22.3	19					
22.9	130			22.5	20					
23.1	131		99	22.9	21					
23.4			100	23.5	22					

23.4	132			23.6	23
23.6	133		101	24.1	24
23.9	134		102	24.8	25
24.2	135			25.4	26
24.3			103	25.5	27
24.8	136			26.0	28
					29
24.9			104	26.2	30
25.1	137			26.6	31
25.3			105	27.0	32
25.4	138			27.2	33
25.7	139			27.8	34
					35
26.0	140			28.4	36
26.3	141			29.1	37
26.6	142			29.8	38
27.0	143			30.5	39
27.4	144			31.3	40
					41
27.7	145			31.9	42
30.0	146			32.6	43
30.3	147			33.3	44
30.7	148			34.0	45
31.0	149			34.7	46
					47
31.4	150			35.4	48
					49
					50
				Total ..	

COMPREHENSIVE RECORD SHEET

Tchr. (or Prin., or Supt.) _____

[illegible]

TABLE 21—Continued

53	54	55	56	57	53	54	55	56	57	53	54	55	56	57
CORRECT ANSWERS					CORRECT ANSWERS					CORRECT ANSWERS				
ITEM	INTELLIGENCE	BACK-GROUND	ACHIEVEMENT	PRACTICES	ITEM	INTELLIGENCE	BACK-GROUND	ACHIEVEMENT	PRACTICES	ITEM	INTELLIGENCE	BACK-GROUND	ACHIEVEMENT	PRACTICES
1	4	mos	1		51	5	0	1 to		101	5	usu	4 or	no
2	5	sel	c	no	52	2	20 o	a		102	4	1 or	Bob	
3	3	sel	b		53	3	0 to	nev	no	103	4	mot	Wil	no
4	2	8 or	Ned		54	4	0 to	a	no	104	1	Eng	Rox	no
5	2	usu	sel		55	1	yes	a		105	4	usu	d	
6	4	sel	5 or	no	56	4	2 or	c		106	3	oft		
7	5	sel	sel		57	1	2 or	b		107	2	gro		
8	1	nev	yes		58	1	3 or	b		108	1	sel		
9	4	0	sto		59	3	3 or	yes		109	1	sel		
10	3	nev	no		60	1	2	no		110	2	usu		
11	2	sel	no		61	2	2 or	b		111	3	b		
12	3	no	b		62	2	mo r	d		112	5	alw		
13	2	3 or	d		63	5	2 or	c		113	2	usu		
14	4	0	c		64	3	0	a		114	5	2 or		
15	5	bot	car		65	1	0	c		115	2	yes		
16	4	3 or	ne l		66	3	0	b		116	5	no		
17	4	yes	no		67	5	1 or	Bob	no	117	2	yes		
18	3	yes	10		68	3	yes	a		118	4	½ or		
19	1	yes	10		69	5	1 or	6		119	2	day		
20	1	yes	3		70	1	yes	b		120	1	Eng		
21	3	no	no	no	71	5	yes	d		121	4	usu		
22	3	1 or	1		72	2	no	d		122	5	oft		

23	2	3 or 4 or bot	a 0 5	no	73 74 75	1 4 1	yes yes nev	b a d	no	123 124 125	1 4 2	gro nev nev		
26	3	bot	12		76	3	0	b		126	1	usu b		
27	1	bot	50		77	5	yes	som		127	1	alw		
28	3	bot	32		78	3	mos	0 to		128	2	usu		
29	1	bot	5		79	5	mos	som		129	3	yes		
30	4	all	18		80	3	mos			130	3			
31	4	yes	b		81	2	0	som	no	131	1	sel		
32	4	yes	a	no	82	3	1 or	no		132	3	oft		
33	2	all	b	no	83	5	1 or	no		133	1	yes		
34	2	mos	b		84	2	0	yes		134	1	oft		
35	5	3 or	app		85	2	1 or	yes		135	3	oft		
36	5	3 or	d		86	1	1 or	no	no	136	3	Rox		
37	4	0	Old		87	4	2 or	b		137	3	you		
38	5	all	no	no	88	5	oft	a		138	5	you		
39	2	yes	no		89	3	4 or	d		139	5	2 or		
40	5	sel	yes		90	4	4 or	sel		140	4	0		
41	3	Joh	flo	no	91	2	dai	a		141	4	yes		
42	4	yes	J		92	2	Ros	usu		142	4	1 or		
43	3	yes	K to		93	1	Mar	sel		143	5	3 or		
44	4	100	b		94	2	Geo	oft		144	1	mos		
45	1	usu	c	no	95	4	6 or	sel		145	2	mos		
46	3	qui	som		96	5	bot	oft	no	146	4			
47	1	1 or	se o		97	3	oft	oft	no	147	5			
48	1	1 or	man		98	1	0	oft		148	4			
49	4	yes	use	no	99	4	0	Ros		149	2			
50	3	yes	2		100	2	at h	Joh		150	3			

CHAPTER XVI

HEALTH, DYNAMIC, PERSONALITY, AND MATERIALS TESTS

1. HEALTH TESTS

The *Comprehensive Achievement Test* contains a subtest on physical health and several subtests which might be regarded as tests of mental and emotional health. But in order to emphasize the great importance of this phase of education, there is reproduced in Table 22 the diagnostic and prescriptive chart used in Baltimore.

Former President Hoover secured and made available to the American Child Health Association several hundred thousand dollars to finance a national school health inquiry. The outcomes of this investigation, directed first by McCall and later by Franzen, are summarized in the following six monographs: ¹

- Number I: *Health Education Tests.*
- Number II: *Physical Measures of Growth and Nutrition.*
- Number III: *Public Health Aspects of Dental Decay in Children.*
- Number IV: *Influence of Social and Economic Factors on the Health of the School Child.*
- Number V: *An Evaluation of School Health Procedures.*
- Number VI: *Physical Defects: The Pathway to Correction.*

The last volume reports a study of physical defects among children in New York City, conducted by the Research Division of the American Child Health Association in coöperation with the Department of Health and the Department of Education. The study was supervised by a Special Advisory Committee, and financed by the Metropolitan Life Insurance Company.

The last volume should be read first since it offers the most immediately practical assistance in the scientific measurement and correction of defects in vision, defects in hearing, dental defects, defective nutrition, impaired tonsils, pediculosis, and in-

¹ Published by the American Child Health Association and distributed now by the National Education Association, Washington, D.C.

adequate health awareness. This volume will refer to the other volumes in the series, as required.

The reader has now been given an illustration of diagnostic tests of common communicable diseases and has been referred to sources for tests of physical conditions which not only may predispose pupils toward disease but also may seriously affect their school progress.

While teachers and other members of the school staff should be concerned with the physical health of the child as a whole, teachers regard themselves as peculiarly responsible for making pupils sensitive to health consideration. They will, therefore, be specially interested in one readily available and easily used outcome of the aforementioned school health study, namely the *Health Awareness Test*¹ for grades III through IX.

For a discussion of the numerous measurements made in physical education, the reader is referred to:

Bovard, J. F. and Couzens, F. W., *Tests and Measurements in Physical Education 1861-1925*, Oregon University Press, Eugene, 1917.

Meredith, Howard V., *The Rhythm of Physical Growth*, University of Iowa, Iowa City, 1935.

2. DYNAMIC TESTS

During the next dozen years the greatest growth in measurement is likely to be in the dynamic realms of wants, desires, or purposes. Since substantial achievement lies in the future, it is more important for this book to give the reader criteria for evaluating future tests than it is to expend much space on contemporary tests.

To do this it will be helpful if the reader rereads the first thesis in the first chapter. There he will discover that the dynamic elements in any person's life are his *purposes*. For psychologically every individual is composed of two and only two apparently discrete entities, namely, purposes and mechanisms for helping the individual realize the purposes which inhabit and drive him. By classifying the multitudinous terms which mean more or less the same thing under these two categories the big "buzzing,

¹ Franzen, Raymond, Derryberry, Mayhew, and McCall, William, *Health Awareness Test*, Bureau of Publications, Teachers College, Columbia University, New York, 1937.

TABLE 22

REQUIREMENTS FOR COMMUNICABLE DISEASES

NOTE: This chart has been prepared for the use of health officers, physicians, school authorities, public health nurses, and others, to serve as a ready reference to facts regarding several of the more important communicable diseases and the requirements of the Baltimore City Health Department concerning them.

DISEASE AND INCUBATION PERIOD	COMMON EARLY SIGNS AND SYMPTOMS	METHOD OF INFECTION	ISOLATION AND QUARANTINE					REMARKS
			1 Isolation of Patient	2 Quarantine of Household Contacts			3 Incidental Contacts	
				If Patient and Contacts Remain at Home	Children under 16	Adults		
CHICKENPOX Incubation Period: 11-24 days Usually 13-16 days.	In children the first symptom noticed is usually the rash which at first consists of small blisters that have developed from small pimples. In a day or two crusts form which fall off in about 14 days. The eruption comes out in crops so that there may be pimples, blisters and scabs all within a small area of the skin.	Contact with a previous case of infection believed to be contained in discharges from nose and throat and skin lesions	Until primary lesions on the exposed parts have healed and scabs are off	No Susceptible contacts may go to school each morning by school nurse during incubation period.	No	No	No.	Very contagious. A mild disease and seldom any after-effects. Important because of possible confusion with smallpox. The long and variable period of incubation and the mildness of the disease do not warrant the quarantine even of susceptible contacts
DIPHTHERIA Incubation Period: Minimum less than 1 day; maximum indefinite. Usually 2-5 days	Sore throat is usually the first symptom in patients over a year old. In very small children it is apt to be croup. In the latter there may be no patches showing, but in the former they can always be found	Contact with a previous case or carrier. Discharges from throat and nose contain infection. Unpasteur-	Until two successive cultures from throat and nose at least 24 hours apart contain no	Yes. Until patient is released from isolation and that cultures from throat and nose show	No Provided patient is properly isolated, and that cultures from throat and nose show	No Provided cultures from throat and nose show no diphtheria bacilli.	No.	Very dangerous, both during attack and from after-effects. The younger the child the greater the danger from diphtheria, except in infants under 6 months of age. There is great variation of type

in "typical" cases either on the throat itself, tonsils, or palate—sometimes on all of them. Some cases look like simple tonsillitis. In nasal cases discharge from nose occurs which usually excoriates the upper lip and may be bloody. Diphtheritic croup simply means diphtheria of the larynx. It is an exceedingly dangerous form of the disease.	Illness usually slight. Onset sudden. Lymph nodes in back of neck usually enlarged. Rash often first thing noticed. Cold in head not a prominent symptom. May have fever, sore throat, and eyes may be inflamed. Rash variable; may resemble measles or scarlet fever, or both.	Until recovery.	Contact with a previous case. Discharges from nose and mouth of a patient contain very infection.	diphtheria bacilli	no diphtheria bacilli.	no diphtheria bacilli, unless food or milk handlers or associated with children or subsequently exposed to infection.	No.	No.	and mild cases are often not recognized unless a culture is taken or subsequent paralysis develops. CHILDREN MAY BE PROTECTED OR IMMUNIZED AGAINST DIPHtheria BY TOXOID INOCULATION. The best time to give toxoid is at the age of six months, or as soon thereafter as possible.
GERMAN MEASLES Incubation Period: 12-21 days. Usually 14-18 days.	Illness usually slight. Onset sudden. Lymph nodes in back of neck usually enlarged. Rash often first thing noticed. Cold in head not a prominent symptom. May have fever, sore throat, and eyes may be inflamed. Rash variable; may resemble measles or scarlet fever, or both.	Until recovery.	Contact with a previous case. Discharges from nose and mouth of a patient contain very infection.	Until recovery.	Susceptible contacts may go to school if inspected each morning by school nurse during incubation period.	No.	No.	No.	A mild disease. Reporting required because frequently confused with scarlet fever. The long and variable period of incubation and mildness of the disease do not warrant the quarantine even of susceptible contacts.
MEASLES Incubation Period: 8-10 days from exposure to initial fever; 12-14 days from exposure to appearance of rash.	Begins with fever followed by symptoms like cold in the head, with running nose, sneezing, inflamed and watery eyes and fever. The rash is usually first seen behind the ears, on forehead and face. It is blotchy and usually dusky red in color. The rash usually appears on the third or fourth day but may occur on the first or as late as the seventh day.	Until recovery.	Contact with a previous case. Discharges from nose and mouth of a patient especially in the early days of the disease before the rash appears convey infection.	Until recovery.	No. Susceptible contacts may go to school if inspected each morning by school nurse during incubation period.	No.	No.	No.	Very contagious especially during the first few days before rash appears. Practically every one who has not had disease is susceptible. Because of these facts measles occurs characteristically in epidemics. Efforts to cut these epidemics short through the quarantine of contacts are seldom successful and cause much inconvenience and loss of school time. The primary object should be

TABLE 22—Continued

REQUIREMENTS FOR COMMUNICABLE DISEASES

NOTE: This chart has been prepared for the use of health officers, physicians, school authorities, public health nurses, and others, to serve as a ready reference to facts regarding several of the more important communicable diseases and the requirements of the Baltimore City Health Department concerning them.

DISEASE AND INCUBATION PERIOD	COMMON EARLY SIGNS AND SYMPTOMS	METHOD OF INFECTION	ISOLATION AND QUARANTINE					REMARKS
			1 Isolation of Patient	2 Quarantine of Household Contacts			3 Incidental Contacts	
				If Patient and Contacts Remain at Home	If Patient Goes to Hospital or Contacts Leave Home			
MEASLES— <i>Continued</i>				Children under 16	Adults			the prevention of deaths through adequate medical and nursing care. Parents should be instructed in early symptoms and told to keep the child home if these develop. If found in school the child should be sent home and the health department notified of name and address. Measles is very dangerous to children under 3 years old. School children nearly always recover unless they are in poor physical condition or are not properly cared for during illness.

TABLE 22—Continued

REQUIREMENTS FOR COMMUNICABLE DISEASES

NOTE: This chart has been prepared for the use of health officers, physicians, school authorities, public health nurses, and others, to serve as a ready reference to facts regarding several of the more important communicable diseases and the requirements of the Baltimore City Health Department concerning them.

ISOLATION AND QUARANTINE									
DISEASE AND INCUBATION PERIOD	COMMON EARLY SIGNS AND SYMPTOMS	METHOD OF INFECTION	1	2			3	REMARKS	
			Isolation of Patient	Quarantine of Household Contacts		Incidental Contacts			
				If Patient and Contacts Remain at Home	If Patient Goes to Hospital or Contacts Leave Home				
SCARLET FEVER Incubation Period: 1-8 days Usually 3-4 days.	Onset usually sudden, with headache, fever, sore throat, and often vomiting. Glands (lymph nodes) of neck usually enlarged. Usually within twenty-four hours the rash appears as fine, evenly diffused bright red dots. The rash is seen first on neck and upper part of chest, and lasts 24 hours to 10 days, when it fades and the skin peels in scales, flakes, or even large pieces. May have sore throat without rash (so-called "scarlatina sore throat").	Contact with a previous case or carrier. Discharges from nose and mouth, suppurating glands, or ears of a patient. Unpasteurized milk may convey infection. Often spreads through mild, unrecognized cases.	Until 21 days from date of report and until discharges have ceased, not more than 90 days.	Yes. Until release of patient.	No. Provided patient is properly isolated, unless evidence of infection is shown or unless food or milk handled, or associated with children or subsequently exposed to infection.	Children under 16 Adults Children under 16 Yes, 7 days from date of removal. Adults—No.	No.	Dangerous both during attack and from after-effects. Running ears and discharging nose or suppurating glands greatly prolong the infectious period. Great variation in type of disease. Slight attacks may be as infectious as severe ones. Many mild cases not diagnosed and many concealed. A second attack is rare. Most fatal in children under 10 years. Very mild cases may show no rash and occasionally the peeling may not be noticeable.	

SMALLPOX Incubation Period: 7-21 days Usually 7-11 days from effective exposure to initial fever and 10-14 days from exposure to appearance of rash.	Onset sudden, usually with fever and severe backache. About third day usually upon subsidence of constitutional symptoms there develop red pimples felt below the skin, and seen first about the face and wrists and most on exposed surfaces. They form small blisters and after two days more become filled with yellowish matter. Scabs form which begin to fall off about the fifth day. In mild cases pimples and blisters may closely resemble those found in chickenpox.	Contact with a previous case. Discharges from nose and mouth, and contents of pustules are believed to convey infection.	Until 14 days after onset and until skin has healed. Patient usually required to accept hospitalization at Station or at Sydenham Hospital.	Yes. Until 21 days after isolation of patient has been terminated. Vaccination required. If vaccinated within 3 days after exposure, and if resistance is changed may be released from quarantine as soon as health officer determines vaccination has been successful.	Yes. Until 21 days after removal of vaccine. If vaccinated within 3 days after first exposure, may be released from quarantine as soon as health officer determines vaccination has been successful.	Yes. For 21 days after last exposure. Vaccination required. If vaccinated within 3 days after first exposure, may be released from quarantine as soon as health officer determines vaccination has been successful.	Very contagious. Cases of modified smallpox may be, and often are, so slight as to escape detection. Existence of disease may be concealed. A severe type of infection may result from exposure to a mild case
WHOOPING COUGH Incubation Period: 2-14 days. Usually 5-8 days.	Begins with cough which is worse at night. Symptoms may at first be very mild. Characteristic "whooping" develops in about two weeks, and the spasm of coughing sometimes ends with vomiting. If a child vomits after a hard spell of coughing he probably has whooping cough.	Contact with a previous case. Discharges from nose and mouth especially in the early stages before the whoop begins.	Patient restricted from associating with children or attending public assemblies until 28 days after onset and until recovery, but not more than 8 weeks.	No. Susceptible contacts may go to school if inspected each morning by school nurse during incubation period.	No.	No.	After-effects often very severe and disease causes great debility. Relapses are apt to occur. Second attack rare. Specially infectious for the "whoop" two before the "whoop" occurs. Great variation in type of disease. Often fatal in young children, and the weak and aged. Nearly half the deaths are in children less than a year old and 95 per cent are in children under 5 years old

Issued by Bureau of Communicable Diseases
DAVID H. ANDREW, M.D., Director
HUNTINGTON WILLIAMS, M.D., Commissioner of Health

blooming confusion" resulting from terminological redundancies will disappear.

Thus, knowledges, skills, abilities, powers, and mentality are successively more integrated *mechanisms* which the individual utilizes to consummate his purposes.

Thus wants, desires, drives, urges, ideals, attitudes, wishes, appreciations, motivation, interests, readiness of neurones, attention, concentration, persistence, character, moral development, socialization, and "personality patterns" are just so many other words for purposes. Strong and abiding purposes, wants, desires, ideals, attitudes, interests, drives, and the like, automatically compel motivation, interest, attention, concentration, persistence, and the like, so we need pay no regard in measurement to the latter items except, perhaps, as indices of the former items. Commendable purposes are the equivalent of good character, proper moral development, and satisfactory socialization, since the chief ingredient in good character, as usually thought of, is the quality of the purposes and not the efficiency of the mechanisms. A personality pattern such as introversion, for example, simply means that the individual desires or purposes the good opinion of others so strongly that he is easily hurt or purposes to dwell in solitude in preference to extroverting himself in company. But since personality is best defined as the sum total of an individual or his total impact on others, and since this is a composite of both purposes and mechanisms, it is better to call personality tests only those tests which stress both purposes and mechanisms.

The statement that good character is largely or wholly a matter of desires or purposes is so important as to merit amplification. Perhaps a modern legend will clarify the matter:

God first made a woman and thought a long time about how to make her happy. He decided the best way was to equip her with many desires, the gratification of each of which would cause her to feel happiness. Then He created a second woman and in His infinite love (some say) or through error of judgment (say others), He endowed her with a similar set of desires. So long as there was but one woman, there was no problem of character or morality. But when two women appeared with duplicate desires, trouble began. Their overlapping desires when there was a shortage of supplies resulted in unfortunate character manifes-

tations. God pondered the problem. He considered but rejected the idea of providing each woman with entirely dissimilar desires. To do that would mean the practical elimination of moments of happy companionship, and it would seriously restrict the number of desires and hence amount of happiness each might have, since He planned to have millions of persons on the earth. He considered and adopted a very ingenious plan. He strengthened the desire for companionship. He injected in each the new desire to have the approval of the other. He even succeeded in edging-in a feeble desire to give way to the desires of the other. Lastly, He endowed each with intelligence wherewith to eliminate shortages in supplies and to devise optimum methods of accommodation of conflicting desires until the day should arrive, if ever, when there would be no shortage.

Emotions remain to be explained. Some psychologists feel rather shaky about the emotions and hence ignore them; some are so shaken by them that they become incoherent; some seem to find them deleterious to learning and therefore dismiss them with an unceremonious condemnation in general; some with an ascetic turn of mind class them along with Freudianism as not being very nice; and some give them a sort of mythical significance, believing emotions to have a peculiar psychology all their own. The rank and file of us who do not pretend to any abstruse knowledge of psychology occasionally regret that an emotion sometimes disturbs the clarity of our thought, but we do not hesitate to prefer our emotions to the lucidity of a Newton. And, in this, the rank and file of us are right.

This "big, blooming, buzzing confusion" at the very mention of emotions is due to the fundamental misconception that they are few in number, and that all of them are charged with dynamite. Just as truly as there are emotions of anger and of love, there are emotions for triangles, binomial theorems, apple trees, and coefficients of correlation. Emotions are not few; they are legion. They are not all powerful; some are very weak. We have made the grievous error of labeling as emotion only those feelings that are powerful, and we have thus lost sight of their fundamental psychological continuity with the multitude of feeling states which constitute the mass of our affective life. The same psychologist who condemns emotions as deleterious to learning pleads for the value of interest, which is really a kind of emotion

or feeling. The truth is that the worth of an emotion to learning depends upon its relevance. Given continuous relevance, the stronger it is the better it is.

The degree of an emotion is exactly equivalent to the *intensity* of *desire* or purpose, and the object or direction of emotion is identical with the direction of *desire*. In sum, there is no need to carry around with us an extra psychology of the emotions.

Thus we arrive at the thesis that the psychology and measurement of a multitude of supposed entities, reduces to just two.

Even with the foregoing simplification, the matter is complicated enough, for purposes vary in several ways, and each needs to be measured.

Purposes vary in *number* or *variety*. Some pupils are rich in purposes; others are barren.

Purposes vary in *kind*. There are purposes *to do* certain specific acts such as play "hooky" and go swimming. There are purposes *to be* a certain sort of person, such as a "jolly good fellow" or an athletic hero. There are purposes *to enjoy* or appreciate art, literature, music, and the like. There are purposes *to believe* certain theories, dogmas, and the like.

Purposes vary in *intensity* or amount. Some are so weak that they are undetectable once the pupil is out of sight of his teacher. Some are so strong that they make satellites of many of the pupil's other purposes. Possibly the amount of a purpose and the intensity are really different. A purpose may function frequently or invariably without being very violent or intense.

Purposes vary in *permanence*. Some last a lifetime, whereas others, even intense ones, are quite ephemeral.

Purposes vary in their *origin*. Some are intrinsic and others are extrinsic or derived from intrinsic ones. Thus the intrinsic desire for approval or to avoid pain may generate many secondary purposes, such as the purpose to brush hair and wash behind the ears.

Purposes vary in *worth*. Some are deemed to be of great moment by society or by the individual. Others are considered trivial. Worth alters in peculiar ways with the amount of the purpose. Thus, truth-telling is worth more and more up to a certain point beyond which it becomes a positive nuisance. Obedience beyond a certain point invites tyranny.

Purposes vary in *adjustment*. All persons are, of course,

equally selfish in the sense that they always follow their own urges, but that person is labeled selfish whose purposes do not harmonize with the purposes of others.

Purposes vary in *acceptability*, i.e., in the extent to which they are present in others. This, like the preceding aspects, is a highly significant matter to the teacher. It is safe enough to teach mechanisms, for they are inert unless called into action by a purpose. The ability to extract the square root of a number or say an acceptable grace at meals may lie dormant for a lifetime. But not so intrinsic purposes. Purposes are of the heart, and "out of the heart are the issues of life." All the dynamic aspects of life are in purposes. They are dynamite! The teacher who misjudges the acceptability of a purpose or deliberately disregards it, and inculcates a purpose that is highly unacceptable, not only engages in indefensible indoctrination but will probably lose her job besides. The opposite error may have equally serious consequences.

Purposes vary in *integration*. The pupil not only has the problem of adjusting his purposes to the purposes of others but also the task of adjusting some of his own purposes to some of his other purposes. In short, he has the task of lifting to the level of consciousness conflicting beliefs and warring desires, and, through critical examination, achieving an integration among them. Some over-zealous persons contend that, since every integration is unique, measurement is even more impossible in this area than in others. Still more zealous persons go so far as to hold that all integration patterns are equally desirable and that, at most, all that measurement ought to do is to reveal the pattern and refrain from scoring it as good, bad, or indifferent, and also all that teachers ought to do is to discover the pattern and aid it to proliferate, refraining from attempts to alter the pattern.

These are the chief criteria by which educators may test the adequacy of the tests of purposes that will be proposed with increasing frequency. They also provide criteria for judging the teaching of purposes.

How measure purposes? One way is to ask the pupil to introspect and report over his signature concerning whether he possesses certain specified purposes, how strongly he holds them, how consistent they are with other beliefs, *et cetera*. This as-

sumes honesty of report. The assumption is generally justified unless the pupil has special reasons for reporting dishonestly. Some items in the *Comprehensive Achievement Test* are of this type.

When the examiner wishes information for a group, he can neutralize the temptation to report dishonestly by asking for an anonymous report. But anonymity tempts to carelessness, which may be worse than some dishonesty.

Another method is to conceal the test under an innocuous title and disguise the items. Many items in the *Comprehensive Achievement Test* are of this sort. The reader might in all confidence say how many radicals are in this list: Franklin Roosevelt, Benito Mussolini, Calvin Coolidge, Joseph Stalin, Herbert Hoover, Norman Thomas, Karl Marx, Henry Ford, Leon Trotsky, John Lewis, and Nickolai Lenin, without once suspecting that perhaps the more persons he names the more conservative he reveals himself to be. This test item may be taken as a sample of all the tests which use the technique of the razor-edge balance.

Another method is to observe the pupil's behavior in situations where the mechanisms are surely adequate and failure to behave in a defined way can be clearly ascribed to absence of the purpose.

The *Comprehensive Achievement Test* contains items which test a little of everything. It tests some knowledge, some skills, some general methods, some purposes, but its tests of purposes need to be supplemented by observations of behavior. Consequently, in the experiment with activity teaching in the schools of New York City, Wrightstone, who has had most experience with this type of measurement, developed codes of observation, some of which were designed mainly to measure purposes. His book, *Appraisal of Experimental High School Practices*,¹ will acquaint the reader with the operation of his codes.

Since the purposes which exhibit themselves in school under the teacher's eye may be a form of protective coloration, even these codes need to be validated by applying them in situations where the pupils are under the, to quote St. Paul, "law of liberty" in a wide sampling of situations. The writer believes that in time we shall devise pencil-and-paper tests that measure pupils' purposes with sufficient validity and reliability.

¹ Bureau of Publications, Teachers College, Columbia University, New York, 1936.

3. PERSONALITY TESTS

Rating Scales.—The method of rating by means of some sort of scale is the most common technique employed for measuring personality in whole or in its aspects.

Rugg made an exhaustive study of ratings and came to the conclusion that they were so inaccurate as to be practically worthless. The writer, on the contrary, contends that there are few if any measurements made by any science that are more accurate than ratings on personality traits. Rugg is probably right if we inquire whether a high rating for intelligence means that the person rated really has high intelligence. But it is possible that how much intelligence people think an individual has is of greater moment than how much he really has. In this vitally significant area, ratings are delicately accurate.

They are even more significant in the case of personality traits that have no existence outside the mind of the rater. Thus, an individual's force, kindliness, tact, and beauty are in very essence a matter of the subjective impression created on others.

So in this sense, and it is a very important sense, subjective measurement is exceedingly accurate. Nor can we condemn the measurement as being inaccurate simply because two raters disagree. They may disagree widely and yet both be perfectly correct. We should not expect perfect agreement, since the two persons doing the rating do not themselves have identical personalities. The logic of ordinary methods of determining reliability does not quite hold.

Hence the technical problems are not those of reliability or validity—these can be assumed—but of sampling and permanence. Thus, we may wish to secure many ratings in order to discover whether the individual impresses all persons of all types, ages, and sexes about the same or quite variously. Or we may wish to discover whether first impressions are or are not enduring ones. A young lady called at the writer's office to ask for assistance in finding a position. The writer had not seen her before. Her first impression—and a sympathetic one, too—was that she lacked sufficient intelligence for the type of position she desired. Sensing his sympathy she recovered from a shyness not previously visible and conversed with surprising brilliance. The first rating was a correct measure of the impression created. The

revised rating was also a correct measure. The first impression might have proved of considerable importance to her, for professors do not always have an hour available to visit with callers. At the conclusion of the conference she offered to take an intelligence test to help the writer gauge her abilities. The test showed her intelligence to be neither low nor high but average. The writer has perfect confidence in the reliability and validity of both his ratings as indices of her impressions. He is somewhat dubious of both the reliability and validity of the result of the abridged test as a measure of intelligence *per se*, and, of course, even more dubious of it as an index of the impression she produced on him or will produce on others.

Ratings are, then, satisfactory for measuring the impression produced by any aspect of personality, but much less satisfactory for measuring the independent existence of aspects that do not have to depend upon impressions for their sole means of registration. One influence which tends to distort subjective ratings of these independent aspects is the "halo effect," exhibited most dramatically in the case of lovers. It requires long training in science to keep a lover from over-rating the good traits possessed by the beloved and under-rating the undesirable traits. The cause of science does not appear to be the primary concern of Nature. But the halo effect is not limited to lovers. To a less extent, ratings on separate traits tend to be influenced by the general impression which the rater has of the one being rated.

Many methods of securing ratings have been proposed. There are self-ratings and ratings by others. There are ratings by intimates and ratings by casual acquaintances. There are ratings on scales where the steps are fully defined or barely defined. There are ratings on a man-to-man scale, where each step is defined by writing there the name of some individual esteemed to possess just that amount of the trait.

A Recommended Rating Scale.—But for usual school purposes the technique described in Chapter XII or the grade-score-marking technique described in Book Six are recommended since they yield ratings that may be compared with or combined with other grade or age scores.

The teacher may replace or supplement her ratings by asking the pupils to rate each other. Since pupils find it rather difficult

to rank all the members of the class on one or more personality traits, a simpler procedure is to ask each pupil to write the name of the five pupils who are highest and the five who are lowest on a given trait. By giving a pupil a positive score of one point for each time he is mentioned as being among the highest and a negative score of one point for each time he is mentioned as being among the lowest, each pupil can be given a score on the trait in question. Then the pupils can be assigned grade scores according to either of the procedures indicated above.

Chapter XXVIII suggests a list of traits on which pupils are most constantly rated, though the general trend is toward specific rather than such general traits.

Rating without Embarrassment.—One of the difficulties in securing ratings of an individual from his associates is that those who know him best are likely to be the very ones who are most reluctant to give him an unfavorable rating lest it react to embarrass their friendship. The author has proposed a plan for getting around this difficulty, namely by asking several associates to rate the candidate confidentially, not in terms of what they think about him but, according to what they think others who also know him think about him. The plan and suggestions for remedying personality defects are presented in detail in Creative Experiment 24 in *You And College*.¹

Inter-Trait Rating Scale.—The author has developed a more intricate technique which conceals what is being done so completely that an individual about twelve years of age or older is able to rate himself or be rated by his intimates without embarrassment to either. It is known as the *Inter-Trait Rating Scale*. The procedure is to compare in turn each of a series of traits with some trait that is objectively measurable, for example, intelligence.

While writing the preceding paragraph two of the author's close friends entered his office without ceremony. To keep them quiet, he explained the *Inter-Trait Rating Scale* and asked them to operate the scale on him, pooling their judgments. The result is presented in Table 23 partly to make the procedure clear and partly to amuse the reader.

They thought the author's accuracy was less (—) than his intelligence and were 40 per cent sure of this.

¹ Harcourt, Brace and Co., New York, 1936.

Since they could not rate him down in accuracy without rating him up in intelligence or up in adaptability without rating him down in intelligence there was no particular embarrassment to them or him in these ratings, although the author does not see himself as others see him at certain points. They were not asked to state whether the author was very dull or very intelligent or very accurate or inaccurate, nor even to state how much difference there is between his intelligence and his accuracy.

The astute reader will observe that the size of the per cent of certainty is used as an index of the size of the difference. It appears to serve very well for this purpose, although there are instances where it fails to be completely satisfactory. Thus, small differences between some traits appear to be more readily noticed than between other traits. Also, a small per cent of certainty may mean either a small difference or an indication of lack of knowledge of the person being rated. But these errors are of little practical significance since the net social effect is about the same.

Each 100 per cent, whether plus or minus, represents a measurement end-error, since the rater might like to register a larger difference, if the scale used permitted. This difficulty may be overcome by supplementing intelligence as a calibrator with other traits as calibrators which are nearer to the trait rated 100 per cent. This end-error is rarely serious enough and the need for greater accuracy is rarely acute enough to justify a resort to secondary calibrators.

The accuracy quotient of 120 is found by adding one-half of -40 to the author's assumed I.Q. of 140, and similarly for the other personality quotients. Each per cent of certainty is divided by 2 in order that the personality quotients may approximate intelligence quotients in size, range, and interpretation, since the per cents of certainty typically range over about 200 points, from -100 to $+100$, whereas intelligence quotients typically range over about 100 points, from 50 to 150. On the basis of the results from several intelligence tests the author estimates his intelligence quotient to be about 140, although it is difficult to determine with assurance the intelligence quotient of an adult.

Assuming that the 43 traits represent an adequate sampling of the totality of personality (both purposes and mechanisms), and assuming that all are of equal importance, the author's personality quotient is found to be 150. Said one of the callers,

TABLE 23

THE AUTHOR'S PERSONALITY QUOTIENT AS DETERMINED BY THE APPLICATION OF THE INTER-TRAIT RATING SCALE

TRAITS	ABOVE OR BELOW INTELLIGENCE	PER CENT OF CERTAINTY	PERSONALITY QUOTIENTS $\frac{1}{2}$ THE % PLUS I.Q.
Accuracy	-	40	120
Adaptability	+	50	165
Appearance	-	60	110
Cheerfulness	+	40	160
Conscientiousness	0	0	140
Coöperativeness	+	10	145
Courage	+	30	155
Courtesy	+	80	180
Decisiveness	-	10	135
Democracy	+	70	175
Effectiveness	+	10	145
Enthusiasm	+	10	145
Foresight	+	20	150
Generosity	+	90	185
Happiness	+	60	170
Healthiness	-	100	90
Independence	+	70	175
Industriousness	+	80	180
Initiative	+	100	190
Leadership	+	60	170
Likeableness	+	50	165
Loyalty	-	20	130
Open-Mindedness	+	50	165
Orderliness	+	40	160
Originality	+	100	190
Persistence	+	50	165
Pleasing Voice	-	100	90
Poise	0	0	140
Progressiveness	+	20	150
Punctuality	-	90	95
Refinement	+	40	160
Reliability	+	10	145
Self-Confidence	-	10	135
Self-Control	+	70	175
Sense of Humor	+	10	145
Sincerity	0	0	140
Sociability	-	80	100
Sympathy	+	50	165
Tact	+	60	170
Thoroughness	+	50	165
Tolerance	+	80	180
Truthfulness	-	30	125
Vivacity	-	30	125
Average			150

"That's about right, for you are nicer than you are bright!"

Lombardi's Ph.D. dissertation, now nearing completion, is a study of the validity of the *Inter-Trait Rating Scale*. Probably, it will be published by the Bureau of Publications, Teachers College, Columbia University.

Semi-Objective Tests.—An analysis of the *Comprehensive Achievement Test* will reveal many objective or semi-objective items which measure personality.

Downey's *Will Temperament Test* was the first test in this area to receive marked attention. The numerous adverse criticisms of it cannot deny her credit for being an important pioneer, for it is by such criticisms that a science grows.

The *Bernreuter Personality Inventory* was the next test to catch the popular fancy—and to be severely criticized. Just at present Rorschach's *Psychodiagnostic Test* is arousing widespread interest. Here, a series of ink blots are presented in order, and the examiner is asked: *What* is it? His responses are recorded, and scored under certain categories to reveal his personality pattern. Two dissertations, applying this test and discussing other applications of it in this country and abroad have just been completed at Teachers College, Columbia University, and will probably be published soon by the Bureau of Publications at Teachers College.

Hartshorne and May in the elaborate *Character Education Enquiry* have not only created various tests but have done the most fundamental research work in validating instruments of measurement.

Recently Maller has proved to be particularly ingenious in devising effective tests of aspects of personality.

If the reader desires to go more extensively into this subject, he may study the tests listed in Chapter VII, and read:

Kelley, Truman L., *Essential Traits of Mental Life*, The Harvard University Press, Cambridge, Mass., 1935.

Review of Educational Research: Tests of Personality and Character, National Educational Association, Washington, 1932.

Roback, A. A., *A Bibliography of Character and Personality*, The Sci-Art Publishers, Harvard Square, Cambridge, Mass., 1927.

Symonds, Percival M. and Jackson, Claude E., *Measurement of the Personality Adjustment of High School Pupils*, Bureau of Publications, Teachers College, New York, 1935.

4. MATERIALS TESTS

Book Four must perforce deal inadequately with the measurement of school buildings and supplies. Such tests vary all the way from a score card for school buildings by Strayer and Engelhardt to a tiny instrument which measures the degree to which the lighting in a school room is adequate. The following list of score cards for school plants published by the Bureau of Publications, Teachers College, Columbia University, New York, will give some idea of how numerous are the measurements of *things* which help to make the changes we measure in pupils.

SCHOOL BUILDING SCORE CARDS

- Score Card and Standards for Elementary School Buildings.* By GEORGE D. STRAYER and N. L. ENGELHARDT. Standards: 181 pp. Cloth \$1.70. Score Card: sheet 2 pp. 10 cents. [1933]
- Score Card and Standards for High School Buildings.* By GEORGE D. STRAYER and N. L. ENGELHARDT. Standards: 95 pp. Paper \$1.05. Score Card: folder 6 pp. 10 cents. [1924]
- Score Card and Standards for Junior High School Buildings.* By GEORGE D. STRAYER and N. L. ENGELHARDT. Standards: 161 pp. Paper \$1.60. Score Card: folder 6 pp. 10 cents. [1931]
- Score Card and Standards for the Administration Building of a School System.* By GEORGE D. STRAYER, N. L. ENGELHARDT, and W. S. ELSBREE. Standards: 40 pp. Paper 80 cents. Score Card: folder 4 pp. 10 cents. [1927]
- Score Card to Be Used in the Selection of School Building Sites.* By N. L. ENGELHARDT. Standards not available. Score Card: folder 4 pp. 10 cents. [1929]
- Score Card for the Physical Plant of Normal Schools and Teachers Colleges.* By E. S. EVENDEN, GEORGE D. STRAYER, and N. L. ENGELHARDT. For Standards see next item. Score Card: folder 4 pp. 10 cents. [1929]
- Score Card and Standards for College Buildings.* By E. S. EVENDEN, GEORGE D. STRAYER, and N. L. ENGELHARDT. Standards in press. Score Card: folder 4 pp. 10 cents. [1929]
- Campus Score Card and Standards for Country Day and Boarding Schools.* By GEORGE D. STRAYER, N. L. ENGELHARDT, and THOMAS C. BURTON. Standards: 51 pp. Paper \$1.05. Score Card: folder 4 pp. 10 cents. [1930]
- Score Card of Village or Rural School Buildings of Four Teachers or Less.* By GEORGE D. STRAYER and N. L. ENGELHARDT. Standards out of print. Score card: folder 4 pp. 10 cents. [1920]

BOOK FIVE

GUIDANCE AND EVALUATION OF TEACHING
BY MEASUREMENT

CHAPTER XVII
SUBJECTIVE MEASUREMENT OF THE
TEACHING PROCESS

1. CRITERIA OF THE CURRICULUM

It was pointed out in the first thesis in the first chapter that the proper criterion from which to derive all criteria—even those for measuring the teaching process—is the ultimate criterion—happiness. This led the author in an address to the New York Principals' Association to present the following criteria for guiding and evaluating education. If critical discussion and experimentation sustain these propositions, the school of the future will differ greatly from most schools of the present.

Thesis 1.—The objectives of education should be the same as the objectives of each and every person's life, namely to increase the quantity of human happiness and satisfaction.

Thesis 2.—Therefore, the main objective of education should be to increase the quantity of children's happiness, for pupils are really people and they spend a goodly portion of their lives in school and many die before graduation. Since there are adults in a community, the happiness of children cannot be the *sole* aim of education. The discipline from the distasteful is much less educative than the discipline of self-direction with an acceptance of responsibility for the consequences of decision.

Thesis 3.—Pupils are happiest when they are realizing their own present, uncompelled, wise, worthy, and strong purposes.

Thesis 4.—So long as the purpose is the pupil's own purpose, it is not essential that he originate it. It may be suggested by home, community, the school environment, another pupil, or the teacher. Here is the teacher's opportunity for guidance of the pupil. Here, too, the teacher is in danger of forgetting that *Thesis 4* is subordinate to *Thesis 3*.

Thesis 5.—The pupil's purpose should be a wise and worthy one, for others have a right to ask a minimum infringement on their own legitimate purposes, and in the choice of which of several purposes to pursue, the pupil should be led to consider both

the happiness resulting from a moment and the happiness resulting from the consequences of the moment. But this is a decision for the pupils, not the teacher, to make, except for those few decisions necessary to protect life and those pupils who are unable by their own efforts to secure minimum justice.

Thesis 6.—The purposes of adults are not criteria for evaluating children's purposes. Adult purposes are different purposes rather than better purposes. We grow old, regretting the loss of a world that was.

Thesis 7.—Since the present purposes of pupils are much more likely to be concerned with the present and immediate future than with the past, education should be mainly a frontier enterprise.

Thesis 8.—The past should be regarded as the continuous servant of the present—as a means only—unless pupil purposes point toward it. Education should concern itself with the living and not the dead—with the dynamic and not the static.

Thesis 9.—Knowledges, skills and all such inert subject matter should be regarded as means and not as ends, much as we regard books and baseball bats.

Thesis 10.—Such subject matter, being essential to the realization of a pupil's purposes, should be left to take care of itself. Hence the course of study should be concerned with purposes mainly rather than knowledges and skill mainly. Possibly it should be concerned with purposes solely.

Thesis 11.—A pupil's purpose is more likely to be strong if it has a not-too-far-distant culmination continuously visible to him.

Thesis 12.—A pupil's purpose is more likely to be strong if its realization is some immediately useful social product.

Thesis 13.—A pupil's purpose is more likely to be a worthy one if it is realized through cooperative activity.

Thesis 14.—The best way to help a pupil realize his future purposes is to help him realize better, and criticize more discerningly his present purposes.

Thesis 15.—The best way to provide for a pupil's future is not to aim to give him a mastery of much knowledge and many skills, important though these be, but to help him grow a rich set of purposes.

Thesis 16.—*Theses 2 through 15* are sound only if they are in harmony with *Thesis 1*.

There are 16 criteria given above. Many others will be listed in this chapter. A teacher cannot keep so many criteria in mind when in the presence of 40 pupils. Some simplification is imperative. For practical purposes, the dynamic core of the foregoing criteria is *Theses 3 and 4*. With these satisfied, the other 13 will tend to be satisfied.

The foregoing criteria may be compared with and supplemented by the following slightly less radical statement of criteria prepared by Otis, Morrissett, and the author, and incorporated in the report of the Yonkers' Advisory Committee on the Revision of the Curriculum of the Secondary Schools:

The value of a unit of activity resides perhaps fully as much in the method of presentation or manner in which the pupils engage in it as in the nature of the problem or the area of experience in which the project is carried on. Therefore we may almost say that a unit can not be judged in advance but only while it is in use, or after it has been tried, and the verdict then relates to the unit tried in the manner in which it is being tried, or was tried. The same "Topic" or "problem" or "project" presented by different teachers or in different ways, or to groups of pupils of different mental abilities or different kinds of previous experience may turn out to be entirely different and of entirely different value.

The following criteria of a unit are therefore stated as though applying to a unit while it is being tried. When applied to proposed units, such units should be judged of course in the light of *probability* that they will fulfill the qualifications if conducted in the proposed manner. Each criterion implies that *other things being equal* that unit about which the answer is *yes* to the criterion question is a better one than one about which the answer is *no*.

It will be observed that certain of the criteria apply more particularly to the type of unit where the activity is a means to some satisfying end than to the type of unit where the activity is satisfying in and of itself. However, each one doubtless applies to either type to a greater or less extent.

1. When the unit is engaged in are the pupils interested in the activity and/or do they accept the purpose of the unit as their own purpose? And to what extent? (Both as to intensity and intelligence of purpose and number of pupils purposing.)

2. Is the unit planned to culminate in a focus, consciously and continuously present, whose realization is possible and which serves as a criterion of the relevancy of what is done? (The preparation for a concert by an orchestra gives focus to the practice.)

(3) The study of the problem meets the criterion of social service once the pupils could inform their families of their findings and later use their own knowledge when voting.

(4) The unit has ramifications. It calls for seeking sources of information and collecting materials, interviewing public officials, studying various types of city governments, writing letters to obtain data on the experience of other cities, evaluation of arguments, study of organization of private business, study of methods of efficiency, and finally the making of a decision.

(5) The unit meets the criterion of research at the frontier of human decision, for citizens everywhere are now giving serious thought to the question.

(6) It involves coöperation and affords ample opportunity for participation by pupils of varying abilities.

(7) If properly conducted, the unit would give excellent training in the technique of research, in critical evaluation and unbiased judgment. This calls for skillful guidance on the part of the teacher and the pupils must be allowed opportunity to take responsibility themselves.

(8) Whether or not the unit tended to inspire right ideals would probably depend largely on the inspiration of the teacher and those with whom the pupils came in contact in conducting the study.

(9) It meets the rounding-out criterion if the curriculum is not already provided with problems of this kind.

(10) Whether it contributes to some larger integration—a dynamic integration around pupils' purposes and not a superficial, verbal, integration only—depends upon whether pupils have experienced or will experience psychologically allied activities.

The foregoing criteria for judging instruction, units, and courses of study assume an acceptance of the principles of activity education. Some educators may desire subjective criteria, which, while forward-looking, do not represent such a wide departure from customary methods of teaching and courses of study.

Bruner, Stratemeyer, and their students have developed a very elaborate set of criteria for judging courses of study, and have performed the herculean labor of evaluating by means of their criteria about 10,000 courses of study. The criteria have been modified from year to year. The latest form,¹ presented below, may be used to evaluate a course of study or portions thereof or to guide the preparation of a course of study:

The criteria have been divided into four parts: Philosophy, Content, Activities, and Evaluation of Pupils' Work. Under each of these

¹ *Teachers College Record*, November, 1937, Bureau of Publication, Teachers' College, New York.

sections appear sub-sections lettered A, B, C, etc. The criteria themselves are listed with Arabic numerals under the sub-sections. In the sub-sections lettered A, B, C, etc., an attempt has been made to state in condensed form the general ideas involved. The separate criteria under Arabic numerals are intended to define more clearly different aspects of the area under consideration. These latter statements are not altogether mutually exclusive although each contains an additional idea. Neither is it assumed that they cover all the possible points which might be subsumed under the general sub-sections, such as social philosophy, educational philosophy, and principles of learning. It is rather the intent to define more clearly to the user of the criteria the type of social philosophy, educational philosophy, or principles of learning the makers of the criteria had in mind. It is believed, however, that a sufficient number of definitive statements have been employed to cover fairly adequately the important ideas that must be held in mind under each sub-section while a particular course of study is being judged.

A gross scale of four points, Excellent, Good, Fair, Poor, and an item, Not in Course (that is, the item to be rated does not appear in the course at all), has been set up for the course as a whole, for each of the four large sections, and also for each of the sub-sections. This scale can be employed (1) by writing in the appropriate symbol, that is, "Ex.," to the left of the statements where Arabic numerals are used and (2) by checking the appropriate step on the scale under the sub-sections, the sections, or for the course of study as a whole, e.g.,
 Ex. .. G. ... ✓ .. F. P. Not in Course ..
 The order of rating is cumulative, that is, the evaluator would rate the items following the Arabic numerals first, then form a composite rating for the various sub-sections, then, in turn, for the four main sections, and, finally, for the entire course. In each case the rating would depend upon the judgment of the evaluator. In the Laboratory a period of training is given to the evaluator before actual ratings are made.

It is obvious that users of the criteria could, if they so desired, use the statements under the Arabic numerals merely for clarifying purposes, and rate the course only for the sub-sections and for the four main sections, emerging with a composite rating for the entire course.

CRITERIA FOR EVALUATING COURSE-OF-STUDY MATERIALS

I. PHILOSOPHY

Ex. G. F. P. Not in Course

A. SOCIAL PHILOSOPHY

The social philosophy should be one which would do most in forwarding the ultimate aims of a liberal democracy. It should recognize the dynamic character of society and should demand that the school be an active conscious agent for social improvement.

Ex. G. F. P. Not in Course

-
- 1. Is the desirable society conceived of as a democracy?
 . 2. Is it recognized that institutions are to be continually modified as new situations demand and as we achieve better insights and understandings?
 3. Is living conceived of as a process of making adequate adjustments to a dynamic world?
 . . 4. Is social life considered necessary for the fullest expression of the individual?
 5. Is there a recognition of the conflicting forces and issues that exist in life, and have provisions been made to deal with them realistically?
 6. Is the school recognized as a conscious agency for social improvement?

B. EDUCATIONAL PHILOSOPHY

The educational philosophy should be based upon the social philosophy and should be the dominating force in determining the character of the subsequent parts of the course of study. The chief aim of education should be to assist individuals to become increasingly self-directive in improving society through satisfying individual growth.

Ex. . G. F. P. . . . Not in Course . . .

- 1. Is the curriculum thought of as including all the activities of pupils, both in and out of school, over which the school exercises a directing influence?
 2. Is significance attached to relationships existing between the pupil and his environment?
 3. Is the aim of education conceived of as the development in individuals of the ability to direct intelligently their own thinking in regard to their betterment and the improvement of society?
 4. Is significance attached to the fact that people are important environmental factors in experience?
 5. Is it recognized that the school should provide adequate opportunities for differentiated education to meet individual differences in attitudes, interests, understandings, abilities, needs, and skills?
 6. Is the course of study considered as a suggestive guide rather than a rigid outline of materials to be taught?

C. PRINCIPLES OF LEARNING

The course of study should be consistently based on the soundest principles of psychology.

Ex. . . G. . F. P. . . . Not in Course

- 1. Is each new learning act considered to be in some degree remaking the whole organism?
 2. Is self-activity considered fundamental to learning?

3. Is study conceived of as an attack upon the situation, "and what is learned is learned as and because it is needed for the control of this situation"?
4. Are provisions made for taking into consideration the underlying principles of integration?
5. Are the activities and materials organized into patterns which, if used, assist in the better growing of the individual?
6. Is the position held that the learner should experience satisfaction from engaging in activities?
7. Is knowledge considered as a means to enable the individual to participate more effectively in life situations?
8. Is significance attached to pupil meanings and insights?
9. Is the view held that growth and learning are continuous throughout the life of the individual?
10. Is provision made for making the situations of the school real and dramatic?

II. CONTENT

Ex. . . . G. . . . F. . . . P. . . . Not in Course . . .

A. AUTHENTICITY

The materials included should be accurate and authentic, based upon the most scholarly findings and concepts.

Ex. . . . G. . . . F. . . . P. . . . Not in Course

1. Are the materials based upon the soundest available primary and secondary source materials?
2. Do the reference materials include or suggest the most reliable primary and secondary sources for teacher and pupil?

B. UTILITY

The materials should be stated in such fashion that they can be utilized in the solution of life problems.

Ex. . . . G. . . . F. . . . P. . . . Not in Course . . .

1. Will thorough understanding of the problems involved be crucial to most of the group using them?
2. Do the materials assist the pupil to develop and foster a more critical sense of discrimination?
3. Are the data sufficient to arouse in the pupil a keen awareness of the need for problem solving?
4. Do the materials help the pupil to see better his relations as a member of the group?
5. Will the materials help to broaden the social interests of the pupil?

C. ADEQUACY AND SIGNIFICANCE

The materials should be adequate and appropriate in the treatment of those areas of human activity which are most significant for the

welfare of society and the growth of the individual at his level of maturity.

- Ex. G. ... F. P. Not in Course
- 1. Are the materials of everyday significance to society?
 - 2. Is the content included in the course selected to meet the individual and social needs of the pupils?
 - 3. Do the materials include the best thought, past and present, on the most significant and common human and social problems?
 - 4. Do the materials help the pupil understand and exercise in a better way his privileges and responsibilities as a member of a group, thus broadening and stimulating his social interests?
 - 5. Are the materials sufficiently challenging to take into account the needs and desires of each individual at the age and intelligence level considered?
 - 6. Are the materials such that they will arouse in the pupils a keen awareness of the need for problem solving?
 - 7. Does the course of study suggest or include a sufficiently wide range of materials which may be useful in the development of problems or areas?
 - 8. Is a sufficiently representative range of significant points of view regarding controversial issues included or suggested?
 - 9. Do the materials provide adequately for the total present experience of the pupil?
 - 10. Does the course of study make adequate provision for the proper use of physical as well as academic materials?
 - 11. Do the materials lend themselves to the securing of intangible outcomes, such as appreciations, attitudes, and certain techniques?
 - 12. Do the materials provide for various types of learning experiences, such as building, reading, and creating?

D. ORGANIZATION

The material should be organized around major areas of experience so that the pupil may be assisted, first, in discovering and developing promising immediate interests, second, in identifying and satisfying those needs which have value, and, third, in securing an enriched experience.

- Ex. ... G. F. P. Not in Course
- 1. Are the materials organized around broad areas of significant human experience?
 - 2. Are the materials developed through the use of a few large and important problems?
 - 3. Is each of the major problems developed through a series of carefully arranged consecutive minor problems?

-
- . . . 4. Are the facts organized around related ideas so that they may help in developing major understandings or generalizations?
 5. Are the materials so organized that the teacher is permitted sufficient latitude in determining the way in which the materials will be used?
 6. Are the materials so organized that provision is made for individual experiences which have worthwhile values apart from the group activities?
 7. Are the materials so organized that provision is made for effective training in information, skills, habits, and desirable attitudes and appreciations?
 8. Are the materials so organized that they lend themselves to optimum use for both teacher and pupil?
 9. Are the materials so organized that provision is made for frequent revision in the light of teacher and pupil evaluations?

III. ACTIVITIES

Ex. . . G. . . F. . . . P. . . . Not in Course ...

A. PUPIL PURPOSING

The activities should provide for the real purposing of the pupil in order to stimulate in him the desire to proceed on his own initiative in planning, in assuming responsibilities, and in controlling to an ever-increasing extent and on continually higher levels (a) what is to be experienced, (b) the process of development, and (c) the evaluation of the results.

Ex. . . G. F. P. . . . Not in Course . . .

- . . . 1. Do the activities provide for real purposing and planning which will stimulate in the pupil a desire to proceed on his own initiative?
 2. Do the activities result from a problem-solving attitude on the part of the pupil?
 . . . 3. Will the activities give opportunity for the pupil to assume responsibility and to control his experiences to an increasing degree?
 4. Do the activities provide for a clarification of pupils' purposeful ideas through various mediums of creative expression, such as language, painting, drawing, modeling, dramatization, etc.?
 ... 5. Do the activities furnish adequate opportunities for practicing and developing valuable work and study habits needed in accomplishing pupil purposes?

B. INTERESTS AND NEEDS

The activities must be directed toward satisfying real needs, based upon promising interests, to the end that optimum growth may take place; hence these activities must be closely related to the present experiences of the pupil.

Ex.	G.	F.	P.	Not in Course
1.				Are the activities so closely related to the pupil's present life that his own interests will become the natural driving force in initiating and carrying the activities through?
2.				Do the activities promote sensitivity on the part of the pupil to significant needs and problems of his own?
3.				Will the activities, if successfully carried through, result in satisfying present interests and needs and also in creating new and still more valuable interests?

C. SOCIAL VALUES

The activities must provide experiences which, through meeting the demands of an ever-changing dynamic society, will help the child to be a more valuable member of that society.

Ex.	G.	F.	P.	Not in Course
1.				Are the activities concerned with persistent problems and areas of high social significance?
2.				Will the activities contribute to the growth and development of ideals, attitudes, appreciations, knowledges, procedures, habits, and skills which are normally used by children in the important activities of life?
3.				Do the activities provide opportunities for valuable social contacts?
4.				Do the activities assist the pupil in realizing to a greater degree the problems and work of others in making life socially effective and happy?
5.				Is provision made for the consideration of the opinions and suggestions of others?
6.				Is provision made for the individual to seek assistance from the social group and for giving assistance to the social group when such help is desired or needed?
7.				Is there an opportunity for experience in leading and following?
8.				Is provision made for raising the level of social behavior?

D. REALITY

Activities should be provided which are selected from real life situations and which are considered interesting and important by the child because he finds in them many opportunities to satisfy his needs.

Ex.	G.	F.	P.	Not in Course
1.				Do the activities arise from real life situations?
2.				Do they produce, as far as possible, actual life situations?
3.				Are the life situations involved in the activities the most realistic that can be chosen and do they provide the greatest promise for growth in things that matter?
4.				Do the activities provide opportunity for the development of the willingness and ability to face life situations realistically?

E. VARIETY

There should be sufficient variety of interesting desirable activities to provide for the kind of individual and social growth implied in the sections above.

Ex. G. F. P. Not in Course ...

- 1. Is there sufficient variety to provide adequately for pupil purposing and planning?
- 2. Is there a sufficient range of activities to provide adequately for the various interests and needs of the group?
- 3. Do the activities involve a sufficient range of significant social values for the members of the group?
- 4. Is there sufficient variety of activities to enable pupils to face realistically the problems involved?

F. APPROACH

The approach to any series of experiences or areas of work should so challenge every member of the group that each has a chosen desire to initiate and carry to its conclusion the projects which the group has planned.

Ex. G. F. P. Not in Course ...

- 1. Do the materials provide a dynamic approach which will lead to further challenging and accomplishing?
- 2. Are the suggested approaches based upon the present needs, interests, and capacities of the group of which the teacher is the guiding member?

G. CULMINATING ACTIVITY

The culminating activity should constitute a method by which the group and each member of the group realizes the purposes which they have set for themselves. In so doing they will relate and put into the most valuable and meaningful patterns the ideas and materials employed during the entire period of work.

Ex. G. F. P. Not in Course ...

- 1. Has the culminating activity been planned by all the members of the group in the early part of the work?
- 2. Does it provide for the optimum and most meaningful use of the activities and materials utilized throughout the work?
- 3. Is it so set up that pupils and teachers would have opportunity to appraise their own ability to understand appreciations and make functional use of the ideas, activities, and facts employed during the work?
- 4. Has it offered optimum opportunities for a sharing of the work according to the interests, needs, and abilities of each member of the group?

IV. EVALUATION OF PUPILS' WORK ¹

Ex. G. F. P. Not in Course

A. PURPOSE

The purpose of evaluation is (a) to satisfy a desire for a more thorough understanding of the individual child, (b) to provide a basis for intelligent and continuous modification of learning procedures to meet individual differences in abilities and needs of pupils, and (c) to determine the extent to which the accepted objectives of education are being realized and achieved.

Ex. G. F. P. .. . Not in Course

- 1. Is the process of evaluation conceived of as an integral part of the learning experience?
- 2. Does it provide optimum opportunities for furthering the growth process of the individual?
- 3. Do the suggestions for evaluating pupils' work indicate the probability that they will contribute constantly to the improvement of educational procedures?
- 4. Do the evaluation procedures contribute to a realization of the extent to which the accepted educational objectives are being achieved?

B. VARIETY

The evaluation process should incorporate a variety of techniques and devices of measurement and should provide for pupil self-evaluation as well as teacher appraisal of pupils' work.

Ex. G. . F. P. Not in Course . . .

- 1. Does the course of study suggest methods whereby the teacher may evaluate the pupils' work in terms of the individual as well as in terms of the group?
- 2. Is provision made for the individual to appraise his own progress in terms of both himself and his group?
- 3. Are various techniques, such as observation, the oral examination, and the written examination utilized in the evaluation process?
- 4. Are various devices of measurement and of recording pupil growth, such as the anecdotal record, the questionnaire, and the self-rating scale, brought into use in the evaluation process?

C. VALIDITY

The validity of any form of evaluation should be determined by (a) the degree to which this evaluation approximates natural situations, (b) the degree to which the individual accepts the need or purpose of evaluation and participates and coöperates in the process, and (c) the degree that the various aspects of behavior are evaluated in

¹ Dr. Hugh B. Wood is chiefly responsible for the section on "Evaluation."

relationship to other aspects of behavior which emerge to form the whole experience.

	Ex.	G.	F.	P.	Not in Course
	1.				Are the evaluation procedures set up in such a way that they become a natural part of an actual learning situation?
	2.				Does the course of study offer suggestions that will lead to the "acceptance" by the pupil of need for evaluation?
...	3.				Are the evaluation procedures such that they not only permit but tend to encourage the wholehearted coöperation of the individual in the evaluation process?
... ..	4.				Is pupil growth measured in terms of the actual maturation levels of the individual at the time the evaluation takes place?
...	5.				Do all devices and techniques of evaluation have a reasonably high reliability?

D. AREAS OF GROWTH

Evaluation of pupil progress should include the measurement of physical, emotional, and social, as well as mental development.

	Ex.	G.	F.	P.	Not in Course
.....	1.				Is provision made for the measurement of basic skills, techniques, and abilities, such as reading, writing, arithmetic, library skills, and expressional techniques?
... ..	2.				Is provision made for the measurement of basic understandings and informations, generalizations, and concepts in social studies, natural science, literature, fine and general arts, and other areas?
.....	3.				Is provision made for the measurement of desirable intellectual traits, such as open-mindedness, clear habits of thinking, insight, and general mental stability?
.....	4.				Is provision made for the measurement of desirable personal traits, such as ambition, integrity, responsibility, and others?
.....	5.				Is provision made for the measurement of desirable social traits, such as coöperativeness, adaptiveness, and social sensitivity?
.....	6.				Is provision made for the measurement of growth in appreciations, attitudes, and ideals in the aesthetic arts, the social and physical sciences?
.....	7.				Is provision made for the measurement of desirable emotional traits which foster emotional stability, such as love, friendliness, sympathy, and good will?
.....	8.				Is provision made for the measurement of many desirable interests in literature, in the social studies, in science, and in the recreational world?
.....	9.				Is provision made for the measurement of desirable physical characteristics, such as good physique, good stature, and general good health?

E. INTERPRETATION

The course of study should provide definite suggestions for interpreting all evaluation data in the light of known limitations and as nearly as possible in terms of the whole organism.

- | | | | | |
|----------|----|----|----|---------------|
| Ex. | G. | F. | P. | Not in Course |
|----------|----|----|----|---------------|
1. Is the "normal" individual conceived of as one who is not average in every phase of his growth, but as one who deviates from the average in many areas of development?
 2. Is provision made for drawing all evaluation data together into an "integrated portrait" of the individual, rather than using separate and minute data to indicate growth?
 3. Are the interpretation and use of the data in consonance with the purpose of evaluation indicated in A above?
 4. Is provision made whereby the pupil grows in the ability to interpret with increasing accuracy the raw data of his own evaluation in light not only of his own personal development but in terms of his social contributions as well?
 5. Are all evaluation procedures, their interpretations and use, continuously appraised and revised in light not only of their own efficacy but of changing educational goals and objectives as well?

Bruner ¹ gives a more personal and progressive statement of views by listing in a subsequent article the following eleven requirements of an elementary school curriculum:

1. The elementary school curriculum must provide abundant opportunities for developing on the proper age and grade level sounder social and economic understandings.
2. The elementary school curriculum must capitalize in an optimum way upon the educative resources afforded by the local communities.
3. The elementary school curriculum should capitalize upon the educative opportunities provided through the actual social experiencing of children.
4. The elementary school curriculum must attempt to provide for real integration in learning.
5. The subject-matter materials in the elementary curriculum must be accurate and authentic.
6. The elementary school curriculum must make better provisions for the discovery and development of individual aptitudes, interests, and creative abilities.
7. The elementary school curriculum must emphasize the development of problem-solving attitudes and techniques among pupils.
8. The elementary school curriculum must find a more appropriate and effective place for drill.

¹ Bruner, Herbert B., "Some Requirements of the Elementary School Curriculum," *Teachers College Record*, January, 1938.

9. The elementary curriculum should: (a) provide opportunities for children to express their own individuality in the arrangement and decoration of the classroom; (b) call for flexibility in the arrangement and use of furniture; (c) encourage the ingenious use of materials.

10. Many phases of the elementary school program must be advanced through carefully planned and executed research.

11. The elementary school curriculum should make sound and varied suggestions to assist pupils and teachers in evaluating their work.

French ¹ makes the following helpful suggestions to anyone who seeks to change a traditional high school:

1. Ask them (faculty and patrons) to list phases or aspects of current living and thinking in which a number of the students exhibit incompetence.

2. Suggest that an effort be made by the school through a committee to lay a plan by which the school would become a greater factor in creating competence in meeting just one of these important situations or problems.

3. Ask the committee to suggest a plan by means of which they think the school might help the students raise their level of ability and willingness to think and act competently in this situation.

4. Ask this committee or another one to arrange this material in what appears to be a good order for use by a teacher and a class whose sole concern is to meet a problem or situation with greater competence than they have hitherto.

5. Select or create at the first of the next semester, a class or classes to which the material may be appropriately presented.

6. Reorganize, eliminate, and add material as experience with its use dictates, and evaluate the results as objectively as possible to guide further efforts.

7. Select another problem or situation and repeat the process.

8. When the curriculum becomes "crowded" as a result of this process, select "topics" or "subjects" which for some groups of students appear to be of least value and drop them from the curriculum for those students.

9. Allocate these new materials in relation to each other from time to time as experience with their use dictates so that they tend to be placed in a good learning order in reference to the ability, need, and interest of the learner and other factors.

The reader who is especially interested in the subjective measurement of education should not fail to consult the writings of Harap, Hopkins, Caswell, and many others.

¹ French, Will, "Toward A New High School Curriculum," *Teachers College Record*, January, 1938.

2. EVALUATION IN TERMS OF EVOLUTIONARY STAGE REACHED

Every teacher knows that it is one thing to have a knowledge of sound educational principles and a very different thing to make them function in the presence of forty diverse and distracting pupil personalities. All the teachers of the nation are engaged in a vast experiment to discover better ways of incorporating these principles in practicable materials and procedures.

The *Teachers' Lesson Unit Series*¹ was established in order that teachers could share with one another those discoveries which promise to make both teaching and learning happier and more effective. The teachers who have contributed lesson units have not written lectures on pedagogical principles in general, thereby evading the most difficult portion of a teacher's problem. Neither have they assumed that they were writing for perfect teachers with perfect supervisors and perfect pupils. Rather they have told as simply as they could exactly how they themselves taught a given unit in an actual situation.

Among the teachers' lesson units submitted by teachers distributed over the nation are units representing all the main philosophies of education and every gradation from the most conservative to the most progressive. Education in transition is here epitomized. Four stages in contemporary evolution are readily disclosed.

Initial Stage.—Here subject matter set out to be learned is learned for its own sake according to a prior organization, usually logical, and the subject matter is confined strictly to some traditional school subject.

Second Stage.—This is usually known as correlation of subjects either by a single teacher or by a group of teachers in a departmental set up, or by a group of teachers according to Hosis's coöperative plan, in which five teachers, say, take over full responsibility for about 200 pupils. This stage differs from the first only in the breadth of the subject matter. Thus a unit on cotton may consider the history of cotton, the geography of cotton, the arithmetic of cotton, the music of cotton, the art of cotton, and so on.

¹ Bureau of Publications, Teachers College, Columbia University, New York.

Third Stage.—Here, for the first time, dynamic drive and dynamic integration appear. Here there is a natural point of departure, namely, some purpose by a pupil or class, and there is a natural culmination, namely, the realization of this purpose if realization is possible. In working toward the realization of the purpose the subject boundaries are discovered to be artificial and are wholly ignored except perhaps by way of summary. The starting purpose may be one that emerged unexpectedly or may be listed as a portion of the regular curriculum with a definite grade location on the probably justifiable assumption that what has proved to be vital and educative to one group of children will be about equally vital to the succeeding group the following year.

Since what individual pupils will energetically purpose is not so predictable as what will interest a group, it is well to have the common coöperative class activity accompanied by a swarm of individual pupil activities. By observing the power of each pupil's activity to win converts from his associates, and the educative value of the activities that become generally dominant, new units can be found for inclusion in the regular curriculum, thus providing a graded, planned, core curriculum which has been grown coöperatively by teachers and pupils and which is never allowed to get stale and musty and bookish.

The third stage in curriculum evolution is usually reached by the way of the second stage, but it would seem wiser to start with dynamic purposes which fall within the boundaries of some subject. Gradually the teachers and pupils will gain in confidence and skill and soon both teachers and pupils will be jumping the boundaries. Shortly the demand will appear that new wine no longer be confined in old bottles.

Fourth Stage.—The extreme radicals in education fall into two groups, those who are politically and economically radical and who wish teachers not only to have very definite radical objectives but also to use the full power of the schools in a conscious effort to reform majority opinion and reconstruct society. Even though we may agree with their objectives we cannot but recognize that they are merely hastening the day when the schools will become the instruments of the party in power and teachers will be forced to indoctrinate the ideology of the dominant party or get out, even as in Italy, Germany, and Russia of

today. In the long run it would be a wiser policy for all teachers to insist firmly that the schools shall forever remain free from deliberate indoctrination by outsiders or *insiders*—that the schools be recognized as one place where any and all controversial issues may be studied without bias by the new generation.

Switzer has expressed the foregoing idea more adequately in the following lecture¹ which she gave to the students in New College, Teachers College. Her ten points may be regarded as criteria for measuring a teacher's handling of a controversial issue:

Let us list the major considerations which must be taken into account in trying to decide whether a teacher should or should not give his views in the classroom on controversial issues, and then estimate which method is more likely to be beneficial or harmful:

1. *It is important that we have the opportunity to bring controversial issues into the classroom.* A policy has become a tradition in this country—a tradition for which there is much to be said—of keeping out of the public schools issues on which the public is seriously divided. Gradually educators have come to the conclusion that this tradition does not permit them to prepare pupils to become effective citizens. If teachers bring these issues into the classroom and then use them as an opportunity for advancing their own predilections, it is almost certain to have the effect of intensifying the traditional attitude. There is a chance of bringing such issues into the classroom provided educators give complete assurance that they will not take advantage of the situation. They must so act that the most suspicious citizen cannot doubt that the teacher had the issue brought into the classroom for the good of the pupils and not for a selfish purpose.

2. *It is important that we avoid dictation by the nation, state, or city of exactly what opinion teachers must inculcate in pupils.* Let us not deceive ourselves into a false sense of security. One by one the nations of the world are restricting the freedom of teachers to express their own opinions and, furthermore, are dismissing them if they do not advocate the views held by the national governments. China, Russia, Germany, Italy have all recently taken command of the school in this way. The likelihood that this tendency will gain ground in the United States will be increased if teachers insist upon their right to use the classroom as a place to advance their own personal views on controversial issues. Such legislation, as the Nunan Bill, in N. Y. State, the several state bills requiring teachers to take oaths of allegiance, the Dunckel Bill, in Michigan, are straws indicating the direction of the wind. It would be difficult to convince our government

¹Quoted by permission.

that a teacher as an individual is more competent than the government to make wise decisions on complex issues. It is important that all educational organizations agree upon a policy on this question, and publicize and win support for this agreement to prevent the government from dictating what must be taught in the schools.

3. *It is important that pupils should come to sound defensible conclusions on controversial issues.* Even though the teachers could insure sound opinions, which we may be permitted to doubt, this does not tell pupils what to think when future problems arise. In the future they must rely upon their own ability to arrive at sound conclusions. As we have seen, this is more likely to be furthered by the teacher refraining from expressing her views. Most of the pupils come from homes where they are taught not so much how to think, as what to think. They have lived most of their life under authority. This has tended not only to keep children from thinking but has tended to induce them to find comfort and a sense of security in leaning upon authority whether it be dogmatic or thoughtful authority and to find genuine discomfort in having to make decisions for themselves. The teacher being somewhat less interested, we trust, in just what views pupils adopt and much more interested in cultivating the power to think and interest in thinking, is the most promising person to wean pupils from this tendency to accept uncritically dogmatic authority of parents or community.

4. *It is important to teach pupils how to think on present problems.* In proportion as a teacher is able to think soundly, the students will respect him and tend to wait for the teacher's opinion instead of thinking to their own conclusion, and in proportion as a teacher is unable to win respect, we do not want him purveying his opinion to students because it is likely to be unworthy of acceptance. So, whichever way we look at it, it seems certain that students are led to learn to think for themselves best and are led to become suspicious of parents who impose views, if the teacher acts as a neutral chairman concerned primarily with the process of pupils' thinking.

5. *It is important to teach pupils how to think on future problems.* Merely providing pupils with teachers' views on present problems, even though their views be sound, does not help materially to deal with problems which will arise when they have passed from under the teachers' tutelage.

6. *It is important that there be some place where pupils may witness and share in a dispassionate consideration of controversial issues.* The school is, by all odds, the most promising place we can find for exemplifying such consideration. (No other important agency seems as peculiarly qualified as the school to do this.) The effectiveness of the demonstration is likely to be greater if the teacher devotes his energies to guiding the thought-process rather than short-circuiting the process. Pupils are more likely to feel that issues are being dealt with in an unprejudiced fashion if the teacher does not express his views.

7. *It is important that pupils be taught how to think on controversial issues.* Psychology has unquestionably demonstrated that pupils cannot be taught how to think if they do not have genuine problems on which to think, nor can they be taught how to think if a parent does the thinking for them, nor will they be motivated to think if the teacher hands them opinions even though it be at the conclusion of the discussions.

8. *It is important that the public remain willing to support by taxation the schools of the nation.* The struggle between various industrial, political, and other groups for the control of American schools is growing keener every year, and this struggle will be intensified in proportion as the schools deal with controversial issues. Certain large American groups are so fundamentally and bitterly divided that it is impossible for a teacher to advocate a particular view without arousing the animosity of some group. A few years ago, there was a disposition of all groups to be liberal in their financial support of education. The recent pressure against appropriation for public education by Chambers of Commerce and other propertied interests is not due entirely to the depression but is due partly to the growing consciousness on the part of these groups that schools are being used by some teachers as centers of biased instruction.

9. *It is important that teachers remain in the teaching profession.* On every hand we have evidence that excellent teachers are being dismissed from education service, overtly or subtly, because they utilize the schools for advancing personal opinions.

10. *It is important, except for imperative reasons, that public school teachers in a democracy do not take advantage of their position to oppose majority sentiment or views of those who provide their positions and pay their salaries.* As has been indicated, instead of reasons being imperative that teachers give their views regarding controversial issues, we have shown that there are as many or more reasons why they should not do so.

The foregoing discussion relates to public education in elementary schools, *particularly*, in secondary schools, *largely*, and in colleges and universities, *partially*.

Most persons think that logic is the only faculty that should be exercised in making decisions on controversial issues, but emotion has its proper and necessary place. The teacher needs to know the correct steps in the process of making decisions before she can train pupils in this process. The only description of this process, known to the author, that has been published to date appears in *You and College*, Harcourt, Brace and Company, New York.

But this book is more concerned with that other group of radicals who insist that teachers have no objectives of any kind

whatsoever and that standard tests and teachers' examinations should be wholly eliminated from the schools since they are concerned largely with measuring the extent to which certain objectives have been attained. This means, of course, that they disapprove of any predetermined curriculum or any lesson planning. They favor only those objectives which emerge from moment to moment or day by day in pupils' minds, for only thus, they insist, can real dynamic, live purposing be secured. They wish to start a natural living process in a rich and real environment and let purposes emerge from it, be criticized and be accepted, rejected, or modified.

Their aim to secure live purposing and pupil-accepted objectives is commendable but it is possible that the plan outlined as the third stage may actually secure keener purposing. Who has not seen listless children seize upon a suggestion made by an experienced adult and follow it with enthusiasm? A curriculum, coöperatively built by children and teachers, which is constantly being revised through trial and error experimentation may light more sparks than can even the best teacher and the most fertile class illuminated by nothing but the contemporary scene. Annoyed by the mustiness of outmoded curricula, irritated by adult objectives foisted upon children, troubled by objectives in the realm of opinion which induce improper indoctrination or by objectives in the realm of information which have no more validity than much more information in the vast encyclopedia of knowledge, irritated by the tendency of teachers to go straight toward their objectives while forcing the pupils' purposes to follow by fear or extrinsic rewards, they have attacked all objectives, even those in the realms of skill and method which are universally deemed important, and have damned anything and everything connected with objectives. In their admirable zeal for an excellent end, they might better be a bit more discriminating and attack the real errors or else first establish that only wholesale assault and battery will suffice.

A group of educational leaders, who strongly oppose judging the educational process from the point of view of what objectives are being attained, suggest, when pressed for some bases for evaluating instruction, that we are justified, perhaps, in providing *varied experiences* for the pupils. Also they tend to approve a determination of whether pupils are engaged in desir-

able processes—processes leading to objectives continuously emerging out of these experiences and processes but not pre-determined by any teacher or curriculum expert. Among the processes suggested by McGaughy, Mossman, Betzner, and Gans are:

Exploring	Enjoying	Recording
Questioning	Communicating	Reporting
Experimenting	Creating	Planning
Investigating	Practicing	Evaluating
Playing	Thinking critically	

Even so, there appears to be no logical escape from the conclusion that in the last analysis these experiences and processes must be evaluated in terms of outcomes or objectives reached. If the final product produced is not satisfying to the patrons or the pupils in later years or both, the experiences and processes leading to this end-result will be rejected.

For a time the author thought he saw a realm in which processes have validity in and of themselves and in which the formulation of objectives is definitely indefensible. This is the realm of reasonable controversy. Here, surely, it would be indefensible for the teacher to guide pupils toward one of two controversial positions. Here, it seems, is one realm in which the process or way of dealing with the controversy is the all-important matter.

Is this really an exception? Certainly the test of the processes is not to be found in the extent to which they inculcate one of the controversial views. But when pupils are having the experience of dealing with controversial issues, there are several *legitimate* objectives which may properly be used to check on the worth of the process. Such objectives are: the ability to clarify and sharpen the issue, the ability to collect and marshal relevant data bearing on the issue, the ability to discuss courteously and not argue heatedly, the ability to make a tentative decision when evidence is insufficient or refuse to make a decision when evidence is absent.

So, while there is grave danger that teachers will be insensitive to important outcomes, or that they will become so preoccupied with a particular outcome as to short-circuit the process at the expense of outcomes not in the focus of attention, or that

they will fail to recognize the great importance of processes, still, there appears to be no logical escape from acknowledging the final primacy of objectives, especially the ultimate objective, even if pragmatically, as may well be true, we get better teaching by measurement of the process.

CHAPTER XVIII

OBJECTIVE MEASUREMENT OF THE TEACHING PROCESS

The criteria developed in the preceding chapter have been supplemented by others and the whole put in objective test form, ready to be applied to pupils, teachers, principals, superintendents, and patrons. This provides the only semi-objective instruments yet devised for the direct measurement of teaching practices. Form I of the *School Practices Questionnaire* is reproduced here. Its thirteen uses were described in Chapter XIV.

NUMBER RIGHT												
	FS	LC	DS	FST	FA	DC	IA	PL	EV	CO		
	1	6	11	16	21	26	31	36	41	46		

NUMBER RIGHT												
	M	CM	XP	B	KS	TM	A	TX	RC	D	H	
	51	56	61	66	71	76	81	86	91	96	101	TOTAL

SCHOOL PRACTICES QUESTIONNAIRE

A TEST OF THE CURRICULUM

FOR GRADE FOUR THROUGH GRADE NINE—FORM I

Name_____ Grade_____ Boy or girl?_____

Teacher_____ Date of test_____ Age: Yrs._____ Mos._____

School_____ City_____

Instructions. Write your name, grade, etc., in the blanks above.

This test will tell how well you remember what has happened *during the last four weeks*. Try to remember just what happened and tell it correctly in your answers. It does not make any difference *where* it happened. If it happened at home or on the street and happened because of your school life, then it counts. But if it did not happen because of your school life, then do not count it in your answer. The test has nothing to do with your marks.

Read Question 1 in the practice questions below. The true answer,

yes, is circled in the column to the right of the question. Read the rest of the questions on this page and circle either *yes* or *no* for every one.

☞ In the last four weeks—

1. Did *you* go to school? (All through this test, *you* means *you yourself*; it does not mean the class.) (yes) no
2. Did *you* usually sit in the same seat most of the day? yes no
3. Did *you* buy anything in a store? yes no
4. Did *you* discuss anything with the teacher? (*Discuss* means talk over.) yes no
5. Did the teacher *usually* decide what you should do? (When a person *decides*, he makes up his mind.) yes no
6. Did *you and* the teacher make a plan together? (A *plan* is a way of making or doing something, which has been thought out beforehand.) yes no

If you do not know how to circle the true answers, raise your hand and ask for help.

On the following pages, circle either *yes* or *no* for every question. Do not skip any questions. If you are not sure, answer the best you can. If you do not know the meaning of some word, ask the teacher. Ask for no other help. Do not spend much time on any one question. Keep going. When you are told to do so, turn this page and begin. Now let us read the instructions a second time.

TO THE EXAMINER: Give this test only after the class has been working together for four weeks or more; given before that it is not valid. Read aloud the instructions while the pupils read them silently. Make sure that every pupil understands. In the practice questions, inspect each child's answer to each question, help the pupils, make sure that every circled *yes* refers to something which happened in the child's school life or because of it, especially in Question 3, and see that every child in the lower grades knows the meanings of *discuss*, *decide*, and *plan*. There is no time limit for the questionnaire.


FS

☞ In the last four weeks—

1. Did *you and* others discuss why some pupil was not doing his work as well as he could? yes no
2. In school, did *you* spend most of your time on such things as history, English, arithmetic or other school subjects? yes no
3. Did *you and* your teacher try to make someone *like* to do his share of the work? yes no
4. Did *you* discuss in class someone who started talking without waiting for someone else to finish? yes no
5. Did *you and* others study how to get for people a chance to earn enough money? yes no


Go on to the next page.

LC

 In the last four weeks—


- | | | |
|---|-----|----|
| 6. Did <i>you</i> spend most of your time at your own desk or table studying books, writing your lessons, and reciting? | yes | no |
| 7. Did <i>you and</i> the teacher work together outside of school with one or more grown persons? | yes | no |
| 8. Did <i>you</i> talk in class about some work that needed to be done with persons outside of school, and then help do it? | yes | no |
| 9. Did <i>you</i> decide with the class on something which the United States ought to do? | yes | no |
| 10. Did <i>you</i> decide with the class on something which the United States and other countries ought to do together? | yes | no |

DS

 In the last four weeks—


- | | | |
|--|-----|----|
| 11. Did <i>you</i> make up your own mind to do something after talking it over with three or more grown persons or asking them some questions? | yes | no |
| 12. Did <i>you and</i> your teacher both try to have every person who was present do some of the talking in class? | yes | no |
| 13. Did <i>you and</i> your teacher both try to keep people from saying again what someone had already said? | yes | no |
| 14. Did <i>you and</i> the class make sure that each one knew what he should do before the class came together again? | yes | no |
| 15. Did <i>you and</i> your class talk over things about which grown persons do not agree? | yes | no |

FST

 In the last four weeks—

- | | | |
|--|-----|----|
| 16. Did <i>you</i> talk over the good and bad points of someone who wished to be chairman or president or any other officer? | yes | no |
| 17. Did <i>you</i> talk in class about better things for the class to do? | yes | no |
| 18. Did <i>you</i> talk in class about whether war does more good or more harm? | yes | no |
| 19. Did <i>you</i> talk in class about things which labor unions ought to do? | yes | no |
| 20. Did <i>you</i> talk in class about the good <i>and</i> the bad points of your student government or town or city government or United States government or any other kind of government? | yes | no |

FA


 In the last four weeks—

- | | | |
|--|-----|----|
| 21. Did <i>you</i> have to recite, from memory, rules, or products of a country, or number tables? | yes | no |
|--|-----|----|

Go on to the next page.

-
22. Did *you* help decide when the class work should be done? yes no
23. Did *you*, for two weeks or more, plan *and* make something of wood or cloth or metal or paper or clay or anything else? (*Plan* means think out beforehand a way of making or doing something.) yes no
24. Did the teacher always let you explain your conduct before judging *you*? yes no
25. In class, did the teacher or anyone else try to keep *you* from having a chance to talk over any subject which you wished to discuss? (*Discuss* means talk over.) yes no

DC

 In the last four weeks—


26. Did *you and* your teacher keep anyone from bossing someone else? yes no
27. Did *you and* your teacher get people to talk things over instead of starting a quarrel or fight or strike? yes no
28. Did *you* talk in class about how people could keep wars from starting? yes no
29. When pupils, or others, did not agree among themselves, did *you* talk with your teacher about what to do? yes no
30. Did *you and* others get two or more persons to agree with each other, who did not agree with each other before? yes no

IA

 In the last four weeks—

31. Did anyone ask *you* to tell the good *and* the bad points of something which the class wished to do? yes no
32. Did *your* work start several times a week by a lesson being given out by the teacher? yes no
33. Did *your* work start several times a week by your teacher telling you exactly what to do and exactly how to do it? yes no
34. Did *you* suggest two or more things, which the class did after you suggested them? yes no
35. Did *you and* the class plan together how to help someone out of trouble, and then help that person out of trouble? yes no

PL

 In the last four weeks—

36. Did your teacher help *you* to make a plan for getting something done? yes no
37. Did your teacher help *you* to learn how to work with others better than you could before? yes no
38. When *you* did your school work, did you usually do the same things in the same way as the other pupils? yes no

Go on to the next page.

-
39. Did *you and* others make a plan for several persons to do some work together? yes no
40. Did *you* help others to make a plan for doing something by writing down things for several pupils to do? yes no

EV

☞ In the last four weeks—

41. Did *you* get marked mostly on how well you recited and how well you did your written work and your tests? yes no
42. Did other children sometimes examine *your* work carefully and make it better? yes no
43. Did *you* sometimes tell others the good *and* the bad points of your own work? yes no
44. Did your teacher suggest that *you* talk over your plan with someone who might be harmed by it? yes no
45. Did the teacher or any class leader *often* say, "Do as I tell you" or "Do it because it is right"? yes no

CO

☞ In the last four weeks—

46. Did *you* do some school work in two or more committees or groups of pupils? yes no
47. Did your teacher help *you* to do something at school which you wished very much to do but which you did not have to do? yes no
48. Did your teacher help *you* with something which you yourself started outside of school? yes no
49. Did your class or your school have a student government or general organization which was *copied* from courts, police, towns, etc.? yes no
50. Did the pupils of your class have much real power to decide some of the important things in the school government or in the general organization? yes no


M

☞ In the last four weeks—

51. Did *you* choose to make anything which other persons used after you made it? yes no
52. Did anyone at school praise *you* for doing a good piece of work? yes no
53. Did *you* work to get a star or picture or any prize for doing something well? yes no
54. Did any pupil or any other person punish or scold *you*? yes no
55. Did the other children listen to *you* in such a way that you felt pleased? yes no


Go on to the next page.

CM

 In the last four weeks—


- | | | |
|--|-----|----|
| 56. Did <i>you</i> work in a committee or group of children chosen from two or more classes? | yes | no |
| 57. Did your teacher talk with <i>you and</i> others about how to choose a committee? | yes | no |
| 58. Did your teacher help <i>you</i> to learn how to work in a committee or group? | yes | no |
| 59. Did <i>you and</i> your teacher both suggest things for a committee to do? | yes | no |
| 60. Did your class talk about how to visit and talk with persons outside of the school? | yes | no |

XP

 In the last four weeks—

- | | | |
|---|-----|----|
| 61. Did <i>you</i> get a letter from a bank or store or magazine or author or official? | yes | no |
| 62. Did <i>you</i> visit a worker or employer or storekeeper or official and find out something from him for a class or committee or any other group? | yes | no |
| 63. Did <i>you</i> get a carpenter or painter or any other workman to work with you and other pupils? | yes | no |
| 64. Did <i>you</i> get an artist or a doctor or any official to work with you and other pupils? | yes | no |
| 65. Did <i>you</i> get help or advice from any person outside of school who knows a great deal about something? | yes | no |

B

 In the last four weeks—

- | | | |
|--|-----|----|
| 66. Did anyone show <i>you</i> how to search in books for the facts which you need? | yes | no |
| 67. Did <i>you</i> read a book which the teacher suggested to you but which you were not required to read? | yes | no |
| 68. Did <i>you</i> use regular textbooks more often than other books? | yes | no |
| 69. Did <i>you</i> use ten or more different books besides the regular textbooks? | yes | no |
| 70. Did <i>you</i> bring one or more books from outside of school for pupils to use? | yes | no |

KS

 In the last four weeks—

- | | | |
|---|-----|----|
| 71. Did <i>you</i> get facts from outside of school several times a week? | yes | no |
| 72. Did <i>you</i> learn any facts by mail from outside of the United States? | yes | no |

Go on to the next page.

-
73. Did *you* use much arithmetic outside of the arithmetic class? yes no
74. Did *you* learn things mostly in order to recite them, to take tests, and to pass? yes no
75. Did *you* use facts and numbers mostly in order to help decide what to do about something? yes no

TM

☞ In the last four weeks—

76. Did *you* help to make a shelf or bookcase or bulletin board or exhibit or collection *to be used*? yes no
77. Did *you* help to make a toy or puppet or marionette or costume *to be used*? yes no
78. Did *you* use clay or cork or cloth or glass or plaster *to help others make or repair something*? (*Repair* means mend.) yes no
79. Did *you* use leather or metal *to help others make or repair something*? yes no
80. Did *you* use hammer or saw or screwdriver or level or vice or clamps or plane or chisel or sandpaper *to help others make or repair something*? yes no

A

☞ In the last four weeks—

81. Did *you usually* paint or draw the same things as the rest of the class? yes no
82. Did *you* help other pupils to make up a song? yes no
83. Did *you* make up a poem in school? yes no
84. Did *you* use art in *much* of your work outside of any class in drawing, painting, music, weaving or in any other art? yes no
85. Did *you* tell the story of things which you did, by drawing or painting or writing a poem or a song or by dancing or doing a pantomime? yes no

TX

☞ In the last four weeks—

86. Did *you* take any tests (not counting this one) on manners or feelings or right and wrong or happiness or games or hobbies or beliefs or opinions? yes no
87. Did *you* take tests *mostly* on history or language or English or other school subjects? yes no
88. Did *you and* others help make a test which *you* took? yes no
89. Did *you and* others together try doing what persons wished in order to find out whether that would make them more friendly? yes no
90. Did your class try paying no attention to someone who was showing off? yes no

Go on to the next page.

RC

 In the last four weeks—

- | | | |
|---|-----|----|
| 91. Did <i>you</i> make any use of advertisements or price catalogues or real grocery bills or real milk bills or real bank checks in school? | yes | no |
| 92. Did <i>you</i> tell the good <i>and</i> bad points of some report made by a pupil? | yes | no |
| 93. Did <i>you</i> tell the good <i>and</i> bad points of something in a newspaper? | yes | no |
| 94. Did <i>you and</i> others plan together what records to make of a trip or excursion or of a story? | yes | no |
| 95. Did <i>you</i> add something to a library or a collection of pictures or clippings or notes or facts or anything else? | yes | no |

D

 In the last four weeks—

- | | | |
|--|-----|----|
| 96. Did <i>you</i> work with others making plans which the student government or the general organization used? | yes | no |
| 97. Did your teacher on <i>most</i> school days decide the important things for <i>you</i> to do at school? | yes | no |
| 98. Did your teacher plan the work for the class on <i>most</i> school days without help from the pupils? | yes | no |
| 99. Did <i>you</i> on <i>most</i> school days help decide the important things in your class? | yes | no |
| 100. Did your class on <i>most</i> school days use plans for its work which were made by <i>you</i> AND the teacher AND others working together? | yes | no |

H

 In the last four weeks—

- | | | |
|---|-----|----|
| 101. Did anyone at school treat <i>you</i> on <i>most</i> school days like a little child younger than you are? | yes | no |
| 102. Did <i>you</i> , on <i>most</i> school days, do more than the teacher required in arithmetic or English or in any other subject? | yes | no |
| 103. Did any person, on <i>most</i> school days, <i>make you</i> do your school work? (<i>Make you</i> means force you.) | yes | no |
| 104. Did any person at school, on <i>most</i> school days, speak to <i>you</i> in an angry voice once a day or more? | yes | no |
| 105. Did everyone treat <i>you</i> on <i>most</i> school days with as much respect as ought to be shown to a grown person? | yes | no |

In inaugurating a teaching process in harmony with the foregoing *School Practices Questionnaire* certain problems will arise which must be solved. The *Questionnaire* contains within itself concrete suggestions of how to solve many of these problems.

Go on to the next page.

Adams ¹ conducted a national inquiry and discovered that the main problems faced by teachers using activity methods were:

1. How to select and develop an activity.
2. Testing the results of the activity.
3. Maintaining a desirable working atmosphere during the unassigned period and at the same time encouraging individual freedom and initiative.
4. Planning a time program which can be effectively used in developing activities.
5. Adjusting an activity to meet the individual differences of the pupils.
6. Finding the proper reading materials to aid the children in solving their activity problems.
7. Securing other materials necessary in the development of an activity.
8. Securing the coöperation of the principal, superintendent, or other administrative officers not entirely in favor of an activity program.
9. Sustaining the interest of the children during the development of an activity program.
10. Developing an activity and at the same time meeting the course of study requirements for the skill and drill aspects of the curriculum.

She lists in her book a dozen or more solutions for each of these problems—solutions which specialists have evaluated as excellent.

¹ Adams, Fay, *The Initiation of an Activity Program into a Public School*, Bureau of Publications, Teachers College, Columbia University, New York, 1934.

CHAPTER XIX

OBJECTIVE MEASUREMENT OF THE EFFECTS
OF THE TEACHING PROCESS

1. GUIDANCE THROUGH INTERPRETATION OF
TEST RECORDS

Let us ask questions of the data in Table 4, and see what guidance will be given by answering them.

1. *Of which pupil should we expect the most in view of the grade he is in?*—As shown by the G grade, the expectation is the same for all pupils, since all pupils are in the same grade.

2. *Of which pupil should we expect the most in view of his chronological age?*—Numbers 8 and 22 are the oldest, each having an expectation of 3.8 as shown by the Ga.

3. *Of which pupil should we expect the highest achievement, in view of his intelligence?*—We expect most of Number 8, whose Gi is 5.8.

4. *Which pupil made the best record in reading?*—Number 4 is the highest with a Gr of 5.3.

5. *Which pupil made the best record in achievement, i.e., on all educational tests combined?*—Number 14 has the highest record with a Ge of 4.9.

6. *In view of his grade, which pupil most exceeds our expectation in arithmetic?* Number 14. We expect 3.0 and he scores a Ga of 5.3.

7. *In view of his age, which pupil most exceeds our expectation in achievement?*—Number 14. We expect 2.1 and he scores 4.9. The excess of 2.8 is the largest in the class.

8. *Which is the brightest pupil?*—Number 7. We expect a 2.1 and he scores a Gi of 5.6. The excess over the G age is 2.5, the largest in the class.

9. *In view of his intelligence, which pupil most exceeds our expectation in spelling?*—Number 5. We expect 3.8 and he makes a Gs of 5.1.

10. *Which pupils most need to stress reading?*—Numbers 6, 17, 21, 22, and 24 are all weak in reading, which is a serious matter.

If they were equally weak in all other abilities, they might not be able to stress reading without loss elsewhere, unless we could count on a contribution to arithmetic and spelling from the enhanced reading ability, which we generally can. But it happens that these pupils are also generally weaker in reading than in other abilities.

11. *Is No. 13 achieving as much as we expect of him in view of his grade?*—Yes. We expect 3.0 and his Ge is 3.7.

12. *Is No. 13 achieving as much as we expect of him in view of his age?*—Yes, brilliantly. We expect 2.2 and he achieves 3.7.

13. *Is No. 13 a bright pupil?*—Yes, very bright. We expect 2.2 and his Gi is 4.8.

14. *Is No. 14 achieving as much as we expect in view of his intelligence?*—Decidedly not. We expect 4.8 and his Ge is only 3.7. This is probably because no one has recognized his capabilities, being deceived by his low age. Or possibly parents and school are restraining him lest he grow mentally too far away from his age group, fearing a social maladjustment.

15. *Which pupils most deserve the commendation of school and parents?*—Numbers 5 and 18, since the excess of Ge over Gi is the the highest for them. Since their advantage over other pupils is not great it may be due to an error in the measurements. Here as elsewhere small differences should be accepted and acted upon but not with great assurance.

16. *Will No. 8 usually get a higher commendation from the school than Nos. 5 and 18?*—Yes, but he will not deserve it. In fact, he is working 0.9 of a grade below what we expect of him. But it was ever so. “To him that hath shall be given and to him that hath not shall be taken away, even that which he hath.” To man propose this test: Which would you rather have, intelligence and poverty, or stupidity and wealth? Practically all men will choose the former. And yet the alternatives offered by school and society are less just than these. The bright pupil gets everything—approval of teachers, excellent reports, praise of parents, the envy of his fellows, and the plaudits of the greater world. When he becomes a man he often uses his intelligence to exploit his less gifted fellows and grow rich out of their poverty and powerful out of their weakness. And whence came his high intelligence? Through some effort of his own? Not at all. It is an outright gift from the social group—an early bonus. Does he

regard this gift as advance payment on the amount to which he is entitled in this world and hence ask less of life than the stupid who did not ask for stupidity? He does not. Can we expect to see children grow up and establish justice in society when they are accustomed daily to see its most fundamental tenets violated in the school?

How pupils' effort or efficiency (F) scores are related to the brightness (B) scores of pupils is shown by the three accompanying rows of figures. The second and third rows employ T, B, and F scale units. It is readily seen that pupils with low brightness scores tend to have high F scores and *vice versa*. The normal B score is 50, just as the normal I.Q. is 100.

TABLE 24
RELATION OF B SCORES AND F SCORES

Total No. of Pupils. . .	156	428	738	565	165
Brightness (B)	Below 35	35-44	45-54	55-64	65 and above
Median Effort score (F) .	59.1	56.4	54.3	51.3	46.2

17. *Which pupils will learn fastest?*—Those whose G_i most exceeds their G age.

18. *Which pupils will learn slowest in relation to their G_i ?*—Those whose G_i most exceeds their G age. These are the brightest pupils and for various reasons it is easier for dull pupils to keep pace with their slow-growing intelligence.

19. *Which pupils should be watched lest they work too hard?*—Those whose G_e exceed their G_i by large amounts, and all others whose health is known to be frail or whose eyes require guarding.

20. *Which pupils should be urged to go to high school and college?*—Those whose G_i or G_e considerably exceeds their G age. Some would advise urging all to go to high school, at least, and alter it so as to make it a profitable place for both slow and bright pupils to be.

21. *Which pupils should be guided into early vocational choices and preparations?*—Assuming the world to be what it is, those whose G_i falls below their G age.

22. *Which pupils should be urged to avoid becoming typists or enter some other routine mechanical occupation?*—Those whose *Gi* or *Ge* greatly exceeds their *G* age. Here, as in the preceding questions, the decision as to what to recommend should, preferably, be based on a cumulative record of more tests of the same traits and also tests of other traits, as well as on non-intellectual factors.

The functions of vocational guidance can be achieved only through (1) a careful survey of the various occupations to determine the constancy of demand for employees, whether the occupation is a seasonal or ephemeral one, the ratio of demand to supply, the monetary rewards, the nature and amount of other types of rewards, the working conditions in the occupation, etc.; (2) a study of the results of such a survey by the pupil, both to aid him to choose his own occupation intelligently and as an important part of his general education; (3) a testing in various ways of the pupil's ability for and interest in each of the occupations; (4) the choice by the pupil with the advice of a vocational counselor, of his vocation; (5) the provision of adequate vocational education; (6) appropriate educational guidance in the light of the chosen vocation; (7) vocational placement at the end of the pupil's educational preparation; and (8) a systematic follow-up of each pupil sent into industry.

A boy of twelve or a youth of twenty stands before some school official enquiring what occupation it would be advisable for him to enter or for which to begin preparation. What must the educator know before he can give wise advice, and how can measurement help in this intensely human situation?

Sound advice requires the educator or vocational counselor to know the general intelligence limits of the various occupations. This means that intelligence tests must be applied to members of representative occupations. Terman has made some progress in the determination of occupational intelligence limits. The overlapping of I.Q.'s for the different occupations is so great that some college students have less intelligence than some hoboes! The median I.Q. more nearly bring out the true facts, namely, that success as a business man or college student requires an I.Q. considerably in excess of that which is typical for hoboes, salesgirls, firemen, policemen, motormen, and conductors.

A War Department bulletin on army mental tests shows the

intellectual level for various occupations as determined by the application of thousands of intelligence tests at the army cantonments. The scores on these tests, for occupations shown, follow:

- 45 to 49—Farmer, laborer, general miner, and teamster.
- 50 to 54—Stationary gas engine man, horse hostler, horseshoer, tailor, general boilermaker, and barber.
- 55 to 59—General carpenter, painter, heavy truck chauffeur, horse trainer, baker, cook, concrete or cement worker, mine drill runner, bricklayer, cobbler, and caterer.
- 60 to 64—General machinist, lathe hand, general blacksmith, brakeman, locomotive fireman, auto chauffeur, telegraph and telephone lineman, butcher, bridge carpenter, railroad conductor, railroad shop mechanic, locomotive engineer.
- 65 to 69—Laundryman, plumber, auto repairman, general pipefitter, auto engine mechanic, auto assembler, general mechanic, tool and gauge maker, stock checker, detective and policeman, toolroom expert, ship carpenter, gunsmith, marine engineman, hand riveter, telephone operator.
- 70 to 74—Truckmaster, farrier, and veterinarian.
- 75 to 79—Receiving clerk, shipping clerk, stockkeeper.
- 80 to 84—General electrician, telegrapher, band musician, concrete construction foreman.
- 85 to 89—Photographer.
- 90 to 94—Railroad clerk.
- 95 to 99—General clerk, filing clerk.
- 100 to 104—Bookkeeper.
- 105 to 109—Mechanical engineer.
- 110 to 114—Mechanical draughtsman.
- 115 to 119—Stenographer, typist, accountant, civil engineer, Y.M.C.A secretaries, medical officers.
- 125 and over—Army chaplains, engineer officers.

The first step is to utilize tests to define the intelligence limits of the various occupations. The second step in vocational guidance is to measure the individual to be guided to determine in which occupation level his intelligence falls. Then the vocational counselor is in a position to tell the pupil the work he is by intelligence fitted to do. The pupil can be informed that his intelligence approximately equals the average of that of individuals who are successfully engaged in, say, ten different occupations. The pupil may, if he chooses, decide for an occupation that is in the next intellectual level above, but he will not do so without being warned that the higher he aims above his natural level the

smaller become his chances of success. Good luck, family pull, the possession of valuable accessory traits, etc., may cause him to "get along" out of his intelligence element, but he should realize that the attempt would be a speculative one.

Such a determination of a pupil's intelligence is not only advantageous to the pupil, it may be very profitable for an employer, particularly if the employer has an opportunity to choose among applicants. Recently an almost physically perfect youth was given an intelligence test by a member of our psychology department. The test showed him to be feeble-minded. Shortly afterward he was employed as a messenger boy by Wanamaker. A package entrusted to him disappeared. Detectives watched the boy and annoyed members of his family for several days. Later the package was found in the store where it had been carelessly dropped. At the end of the first week the boy was paid and dismissed. He lost his money before reaching home. Several other employers discovered their mistake by the same trial-and-error expensive procedure. Neither the boy nor his family nor his employer profited by these experiences.

A great social waste is the vocational exploitation of the unusually gifted. With certain exceptions every employer is competing with other employers to secure the services of the most competent. The employer does not stop to consider whether he can give the gifted individual, whom he is lucky to employ, abundant opportunity to make the greatest social contribution of which he is capable. The country suffers an enormous loss each year because many of its geniuses have been caught by this exploiting system, and placed in relatively non-productive positions. The individual employer can afford this but society can't. Society's aim is to guide no individual into an occupation above his intelligence. Society is equally concerned that great gifts be not frittered away on small jobs. In sum, we want both minimum and maximum intelligence limits for each occupational level. In so far as it can be done without doing too much violence to individual liberty, the social group should guide each individual to the level fixed for him by nature. Only thus can the social group be most efficient, prosperous, and happy.

In time society will recognize its essential organic nature, and then the persons of low and average ability will themselves insist that the able be placed where they can make the greatest con-

tribution for the good of all. The gifted, considering their superior native endowment as part payment for their services, will contribute to the social group without extorting undue monetary rewards from the group which they serve. Vocational guidance through the schools is about the only way to accomplish this great and beneficent task.

Society cannot safely trust its geniuses to find their own way through the industrial maze. Immature occupational preferences frequently lead where there is no turning back.

The much-debated report of Thorndike and Lorge that there is little or no relationship between intelligence and occupational success may mean no more than this: if all grades of intelligence are driven by the wasteful action of circumstances into the same low-level occupation, they succeed about equally well, the dull pupils working hard to hold their jobs and the bright pupils loafing because there is no intellectual challenge in routine work. Put both groups into positions that really demand the strenuous exercise of intelligence and the conclusions from the investigations are likely to be quite different.

For a more extended treatment of this subject the reader is referred to:

Bingham, Walter, *Aptitudes and Aptitude Testing*, Harper and Brothers, New York, 1937.

Hull, Clark, *Aptitude Testing*, World Book Company, Yonkers-on-Hudson.

Kitson, Harry D., *I Find My Vocation* (Rev. Ed.) McGraw-Hill, New York, 1937.

Morton, Nelson W., *Occupational Abilities*, Oxford University Press, Toronto, 1935.

23. *Is the class over-age or under-age for the grade?*—It is under-age, the G grade being 3.0 and the G age 2.7.

24. *How bright is the class?*—It is very bright. The intelligence is accelerated 1.3 grade, the G age being 2.7 and the Gi 4.0.

25. *In view of the grade, how well did the class do in reading, arithmetic, spelling, and achievement?*—Quite well. It is 0.3 of a grade ahead of the G grade in reading, 0.8 ahead in arithmetic, 0.2 ahead in spelling, and 0.4 ahead in general achievement.

26. *In view of its age, how well did the class do?*—Very well indeed. It exceeded the G age in reading by 0.6 of a grade, in

arithmetic by 1.1, in spelling by 0.5, and in education in general by 0.7.

27. *In view of its intelligence, how well did the class do?*—Quite poorly. The Gi exceeds the reading by 0.7 of a grade, the arithmetic by 0.2, spelling by 0.8, and achievement by 0.6.

28. *How well does the teacher know the abilities of the children and how fairly does she judge them?*—The answer to this question is revealed by the closeness of agreement between Ge and Gt. Since Gt appears in Table 17 instead of Table 4 the question cannot be answered for the pupils in Table 4.

29. *Which comparison is most just to pupil and class—with G grade, G age, or Gi?*—All are meaningful and helpful. The comparison with grade is the poorest since the children in school with practically 100 per cent promotion will be penalized, whereas the school, which fails to promote its pupils readily, profits not only in comparison with the grade norm but also from the reputation made by its maturer graduates when they go to high school.

The G age, i.e., age norm, takes all of the handicap out of easy promotion and all the profit out of retardation, and incidentally shows, when compared with G grade just what is happening. Hence this comparison is fairer.

But, since pupils of the same age vary greatly in intelligence, the comparison with Gi is the most just of the three. But a combined Gi and Gb, i.e., grade score in community background, is fairer still.

30. *How efficient is this school?*—A comparison of class Ge with class Gi indicates inefficiency, but before we may draw this conclusion many things must be considered and are considered in Chapter XXII.

In a city just west of New York City there is an elementary school which has acquired a reputation for its great efficiency. The principal, the teachers, and the pupils are proud of this reputation. Once an invitation was received to visit this school and give some standard tests in order that the efficiency of the school might be revealed in a scientific manner. Both intelligence and educational tests were administered to the children. The tests were scored and the results compared with the norms from a large number of schools throughout the United States. The principal sent word that he was planning a mass meeting of teachers and parents to hear a report of the results of the test.

Because of the increased mental maturity, their students had a distinct advantage over others in high school.

But the report upon the work of the school was not yet finished. The grade for grade comparison shows something about the efficiency of a school, but not enough since children may vary in the age of reaching a given grade in different schools. The age for age comparison shows something also. In fact it shows more than the grade for grade comparison. It is a more delicate measure of efficiency than the other. But there is a more delicate measure than either of these, namely the intelligence for intelligence comparison. Now this particular school was located in the well-to-do residential section of the city. The intelligence test confirmed the guess that the children of these successful parents averaged considerably higher in intelligence than children in general. Since this was so, and since it is much easier to teach gifted children than to teach average or dull children, it is necessary in determining the efficiency of a school to enquire what the school is doing in proportion to the intelligence of its children. When this comparison was made it was found that instead of the school being an efficient one, it was decidedly inferior. In proportion to the intelligence of the children the school should have been doing much better than it was. Thus the school which had acquired a reputation for great educational efficiency, was, by refined methods of measurement, shown to be really inferior in its efficiency. Its reputation had been built up and maintained because it unjustifiably retarded its children, and because it was fortunate enough to be located in a part of the city which sent it very intelligent children. The great sage, Confucius, was wise. He was also clever. He refused to take any pupil for instruction who, when taught three corners of a subject, could not see the other corner himself. This was nothing in the world but a crude intelligence test. Confucius fully realized that the reputation of the teacher depends more on being able to surround himself with gifted students than in being a good teacher.

At the close of the meeting with the principal and teachers, the principal asked that the results of the measurements be kept confidential, that he had been misled by the false reputation of the school into a sense of security, that he would now set energetically about the task of improvement. It was agreed not

to make the results of the test public, but instead to turn them over to the teachers to aid them in the process of developing a real efficiency.

The tests in *A Comprehensive Test Program* described in Chapter XIV permit us to ask and answer many other questions. Chapter XV words these questions and tells how to answer them.

31. *In the light of these findings which subjects should the teacher stress with the class?*—Spelling is most in need of attention and reading next. However, this assumes that the typical emphasis throughout the nation is the proper emphasis. Before deciding finally which subject to emphasize the teacher should reread the discussion of norms as objectives in Chapter XVII.

2. SHOULD TEACHERS REGARD TEST NORMS AS TEACHING OBJECTIVES?

In those areas where it is proper for a teacher to have educational objectives, these goals should:

1. Be visible to teacher and pupil since thereby motivation is markedly enhanced,
2. Be wisely proportioned in the relative amounts of each, so as to regulate emphasis,
3. Be adapted to the teaching power of teachers and the learning capacity of pupils, so as not to discourage effort by demanding too much or too little of both.

There have been various attempts to satisfy these three criteria. The most common method is to publish a curriculum in which the objectives are identified and varying amounts of each allocated to the different grades. No doubt this has value, especially for impressing the public.

Fortunately neither teachers nor pupils take such curricula very seriously in so far as the amount of each objective is concerned, because, if they did, utter discouragement would paralyze effort. The accomplishment demanded by the typical published curriculum exceeds by many times what is at all feasible.

A conscientious principal of a private school in a large city was much perturbed when a program of tests administered by the author revealed that her pupils had not mastered, grade by grade, the published curriculum of the city. The author comforted her with the assurance that the published curriculum was

window dressing for the patrons and that tests would surely reveal this to be so.

A group of teachers, noted for their professional zeal, decided to set up their objectives more realistically. Thus, in composition, they selected and published a specimen of composition to show what amount of general merit in a composition they expected of each pupil at the end of the fifth grade. The composition defined and made visible the passing point for the fifth grade. The author had his students in measurement score the composition on the *Nassau Composition Scale* and reported to the teachers that 25 per cent of sophomore college students could not equal in composition ability what they were demanding of their fifth-grade pupils. "But," replied the incredulous supervisor of the teachers, "that specimen of composition was written by one of our fifth-grade pupils." The teachers erred but they had taken an important forward step. They had made their goal visible, and experience would have made it reasonable.

A bright young teacher in that long ago class in measurement decided to make the objective in reading ability both visible and fair. Since there were many forms of the *Thorndike-McCall Reading Scale* she planned to administer one form each month and graph the results. When the author visited her class during the second month of the following school year, she asked one of the pupils to explain the two graphs to me. He explained the first one, showed me a vertical line that represented the median for the class, and pointed out his position just below the median. Said he, "The teacher told me that my job was to jump that line."

"And did you?" I enquired.

He replied with a peculiar mixture of pathos and pride, "I would have, but the line jumped."—as, of course, the class median would, since the other children were trying to jump the ones just in front of them.

The grade norm, i.e., G grade, is like the class median in that it makes the goal visible, and is superior to the class median in that it is probably a better regulator of emphasis, indicating as it does a sort of universal consensus as to how much of one ability is equal to a given amount of another. But the grade norm, even though it sets a goal that is reasonably satisfactory for a roughly typical class, sets a very unsatisfactory objective for individual pupils and for atypical classes,

The age norm, i.e., G age, has all the advantages of the grade norm plus the extra advantages of not being affected by the grouping and promotion practices of a particular school and school system, but it, too, does not set a reasonably fair goal for all pupils, classes, and schools, omitting as it does several important elements.

The intelligence norm, i.e., G_i , almost perfectly satisfies all the criteria, incorporating as it does the influence of grade and age and, in addition, that very important ingredient, inherited learning capacity.

The expectancy norm, i.e. $[(2 G_i \text{ plus } G_b) \div 3]$, is the most adequate general formula the science of education has developed for setting goals that are visible, for regulating emphasis, and for stimulating progress by pupils, classes, grades, schools, and school systems. In a very real sense, this formula locates the norm or objective within the pupil. It is least satisfactory as a regulator of emphasis, since we cannot be sure that the relative amounts of different abilities which have been developed in the past are in optimum proportion. Here, one's philosophy of happiness as developed in Chapter I must give additional guidance. There is need of constant checking, by experiments and critical thinking, to make as sure as is humanly possible that the proportion is modified to fit different types of pupils and different types of environment.

This does not mean that the expectancy norm is equally deficient as a basis for judging achievement, provided as many traits are measured as appear in the *Comprehensive Achievement Test*, for while one's philosophy may favor more of one trait and less of another, the limitations on a pupil's capacity to learn strike a balance.

For a discussion on the high school level of standards as goals, accomplishment ratio, and marks based on standard tests, the reader is referred to:

Symonds, Percival M., *Ability Standards for Standardized Achievement Tests in the High School*, Bureau of Publications, Teachers College, Columbia University, New York.

CHAPTER XX

TESTS AS TEACHING INSTRUMENTS

The Quest for Efficient Methods of Teaching Skills.—Among educators there is the feeling that teachers' lesson units, although they provide excellent initial activities and core activities throughout the learning period, do not yield, as a product of incidental drill, sufficient mastery of the basic skills. They hold more direct drill to be essential.

How shall this drill be provided? What help can measurement give? Several years ago, the model school of a large teachers college decided to make an intensive effort to develop more effective methods of teaching silent reading. A coöperative ¹ investigation showed that teaching a group of teachers the principles of the psychology and pedagogy of reading did not make these teachers more effective teachers of speed and comprehension in reading. This suggests, though it does not definitely prove, that we cannot hope to produce good teachers by providing them with general principles of teaching and trusting them somehow to translate these into effective procedures. The findings in the Doctor of Philosophy dissertation entitled *Measuring Efficiency in Supervision and Teaching* ² point in the same general direction. Thus evidence is accumulating that the gap between principle and practice is much wider than we have generally supposed, and too wide to justify much of the teacher training given in normal schools and teachers colleges.

A principal of a private school divided his teachers into two groups of approximately equal teaching ability. He paired two teachers in each grade from Grade III through Grade VIII. Initial tests were given. Then, unknown to the control teachers, he invited outstanding specialists in the psychology and peda-

¹ McCall, William A., "How Wide Is the Gap between Principle and Practice?" *Teachers College Record*, April, 1936, Bureau of Publications, Teachers College, Columbia University, New York.

² Crabbs, Lelah Mae, *Measuring Efficiency in Supervision and Teaching* Contributions to Education, No. 175. Bureau of Publications, Teachers College, Columbia University, New York, 1925

gogy of reading to hold weekly conferences with the experimental teachers with a view to helping them to be better teachers of reading skill. These conferences were continued throughout the year. At the end of the year final tests were administered. The growth in reading in the experimental classes was calculated and compared with the gains made in the control classes.

The experimental group actually failed to gain as much as the control group. A cynic might conclude, therefore, that the best way to improve teachers in service is to keep them from becoming contaminated by specialists and books by specialists! In view of the crudities in the experiment and the previous excellent training of the teachers, a truer, and certainly a more charitable, conclusion is that specialists are really quite harmless! The writer pondered this discouragingly small gain from so much effort, and determined to cease experimenting with such conventional and fruitless methods of improving teaching efficiency, to discard the time-honored methods, and to invent a new approach to the teaching of reading and other such skills.

The first effort to solve the problem involved the construction of the nine forms of the *Thorndike-McCall Reading Scale*, the invention of reading age and reading quotient comparable to mental age and intelligence quotient, the utilization of these to calculate a pupil's estimated reading age or objective for the end of the year, the monthly measurement of his progress toward that objective, and weekly informal tests. The details of this procedure are given in an article by Dransfield,¹ who has shown that this invention induced marked progress on the part of pupils, unless we are to make unreasonably large allowance for their merely becoming "test wise."

Since monthly and weekly tests yielded such favorable results, it occurred to the writer that results might be even more favorable if the monthly standard tests were supplemented by daily, informal, objectively-scorable, yes-no reading tests based on the pupils' common textbooks in reading, geography, history, and the like.

The writer conducted, in a large New York City public school, an equivalent-groups experiment to test the worth of this idea. Two large classes in each grade from III through VIII were

¹ Dransfield, Edgar, "A Technique for Teaching Silent Reading." *Teachers College Record*, vol. 26, pp. 740-52, May, 1925.

equated. Other customary precautions for equivalent-groups experimentation were observed. The experimental groups made 80 per cent greater progress than the control groups in comprehension and 25 per cent greater progress in speed.

Next the author simplified the procedure and made it even more effective by combining the daily feature of the informal tests and the standard-test feature of the monthly tests, thus providing a daily standard test lesson.

Thus, little by little, the gap is being bridged between principle and practice by embodying principles in functioning procedures.

The acceptance of such multiple-choice test lessons and of multiple-choice tests in general has been partially blocked by the sincere belief of some persons that we should never present wrong forms to pupils.

Those of us who were brought up on the psychology of the revered William James accepted without question his laws of habit formation. One of these, the dictum that no exception should ever be permitted to occur, seemed so perfectly right that few have given it a critical thought. Hence persons sympathetic to standard tests were genuinely troubled when these tests began presenting one correct choice and, say, three incorrect choices.

Those who gave their loyalty to the law and opposed the spread of multiple-choice tests have been greatly perplexed of late to find experiment after experiment proving the superior educational efficiency of such daily practice tests as, say, the *Standard Test Lessons in Reading*. This has forced a reëxamination of James' law. And the law has not been able to stand this critical inspection. For example, we notice for the first time that it is in conflict with the equally respectable law of trial and error. Further, we find it difficult, if not impossible, to regiment any part of education to such an extent as to prevent exceptions. We find the pupil's mind producing both correct and incorrect responses in great profusion. Some come from himself. Others are provided by playmates and parents. The test may present three incorrect responses, whereas, left alone, the pupil's own mind may present thirty. This cannot be prevented. We may even doubt the wisdom of trying to stop it.

Most of these responses are invisible to the teacher. Multi-

tudes of incorrect ones, since they remain uncriticized or unscored, are accepted by the pupil as correct. The modern test or standard test lesson makes these invisible errors visible and thus makes their elimination possible.

A multiple-choice test does not increase the number of incorrect mental responses. It merely brings these responses into the open where the fierce light of publicity can beat upon them.

In short, it pays in growth and development of pupils to measure and to teach by measurement, and it also pays in dollars and cents. Dransfield repeated the experiment conducted by the writer in the New York City schools in another school system and secured the following results:

	<i>Experimental School</i>	<i>Control School</i>
Average Gain in Reading Comprehension (T Score).....	11.5	7.9
Average Gain in Reading Speed (Questions Attempted).	4.5	3.1
Average Gain in Speed of Comprehension (Questions Answered)	5.7	4.0
Average Gain in Reading Accomplishment Ratio	19.8	12.6
Average Intelligence Quotient	95.5	99.9

Tests were likewise given in fundamentals of arithmetic, problems in arithmetic, handwriting, composition, spelling, and geography. The data from these tests led Dransfield to conclude "that the other school work not only had not suffered but had been improved by the special work in reading."

Finally, Dransfield computed the money value of the gain made and found it to be \$3237 for the control school and \$5092 for the experimental school, making a difference of \$1855 in favor of the experimental school and the teaching method used, not counting the probable superior gains made in other subjects, nor the cumulative effects of the added reading ability through the many years to come, nor the increased teaching skill of the teachers available to future pupils.

Pittman, for his Ph.D. dissertation, undertook an elaborate experiment to evaluate a system of zone supervision guided by standard tests. Initial and final tests were given in practically all the subjects of instruction to all the pupils in an experimental rural county and to all the pupils in a presumably

comparable control county. The experiment lasted one year. Again, the schools where the supervision and teaching were guided by standard tests made vastly greater progress than the control schools. The progress made by the pupils in the experimental county was 94.2 per cent greater than that made by the control pupils. Pittman computed the money value of this greater gain and found it to be \$45,102.

Influenced by such evidence as the foregoing, the Board of Education of Rutherford, New Jersey, invited Crabbs to give two days per week as director of measurement for the Rutherford public schools. Even though the program of measurement and supervision which Crabbs undertook was limited by the two days per week schedule, the extra pupil gain in ability in many subjects was calculated to be between \$10,000 and \$15,000 annually, after allowance had been made for the director's salary and the cost of test supplies.

The cumulative evidence from such practical experimentation leads to the inescapable conclusion that *it pays to measure*. Any school or school system that does not provide for one or more persons expert in the technique of testing and methods of teaching based thereon is deliberately choosing an inferior efficiency.

Test Lesson Program for Reading.—Imagine, if you will, the world and its civilization to be exactly as it is this moment, with the exception that every individual therein, barring one, has reverted to savagery. Imagine, further, that this one educated and civilized individual is a teacher, and that this teacher is you who now reads these words. You, being by assumption a teacher, would through force of habit begin frantically to teach the savages, even though your more thoughtful moments lead you to prefer a state of blissful ignorance. We shall allow the savages to have the present available supply of native intelligence. Now, what one thing, if you were compelled to choose but one, would you teach the inhabitants of your primitive world in the hope of lifting them from savagery to civilization as quickly as possible?

We believe that you would elect to teach them how to read. For this reason we give first place to test lessons in reading.

Gates ¹ says: "The power of comprehension of printed words,

¹Gates, Arthur I., "Study of Depth and Rate of Comprehension in Reading by Means of a Practice Experiment," *Journal Of Educational Research*, January, 1923.

among those mechanically able to read, is constantly practiced to its maximum; a maximum determined by the general mental maturity at the time. It may be that practice sufficient to develop maximal power of comprehension is supplied by the ordinary experiences of home and school. Even intensive specific practice produces no further increase in this ability; further progress depends entirely upon growth.

"Increase in the power of comprehension may depend upon the growth of native ability, which like height may be conceived to be determined by inner development independent of environmental factors save a sufficiency of food, exercise—ordinary healthful living—or, it may depend upon knowledge, breadth of information, and experience. In either case specific practice in the tests would be expected to produce little or no improvement."

It is not possible to exaggerate the importance of the foregoing inferences. It means either that the depth or power of comprehension in reading should henceforth be classed with native intelligence as one of the unimprovables, or else it means that it is useless for teachers to attempt to increase this power beyond what is now being attained, or it may mean both of these. An important corollary of the latter is that there is one trait that home and school are teaching with 100 per cent efficiency. Even the satisfaction emanating from the last corollary, were it true, would not compensate for the lugubrious information that we must remain content with the deplorably inadequate power of comprehension now possessed by the upper-grade pupils. Fortunately, we now know that none of the foregoing inferences are true ones, and Gates probably has altered his views, in the light of such experiments as those just described.

In order to show how extensively measurement has developed test lessons, there follows a test lesson program for the teaching of reading from the cradle (almost) to the grave (almost). When two or more publications are given, any one or several may be used. No attempt has been made to list the numerous work books which accompany regular series of readers and are partly or wholly dependent on them.

PRESCHOOL

Teeny Tiny Rimes, Johnson Publishing Co., Richmond, Va.
This is an ultra-simple book which utilizes a test lesson tech-

nique for teaching reading. It may be used by parents or in the kindergarten. The tests are not standardized.

KINDERGARTEN

Pupil Activity Reader, Book 1, Laidlaw Brothers, Chicago.

Each odd page contains typical primer reading material, and each even page contains test lessons based on the odd page. All the early books follow the same scheme. The tests are not standardized.

GRADE I

1. *Pupil Activity Reader*, Book 2.

2. *Picture Story Reading Lessons*, World Book Company, Yonkers-on-Hudson.

The lessons culminate in a picture which enables the teacher to tell at a glance whether the pupil has correctly read and executed directions. The tests are not standardized.

GRADE II

Pupil Activity Reader, Book 3.

GRADE III

1. *Pupil Activity Reader*, Book 4.

2. *Standard Test Lessons in Reading*, Book 2, Bureau of Publications, Teachers College, Columbia University, New York.

All books in this series follow the pattern of Test Lesson 72 given here as a sample. Each pupil in the class reads the selection and chooses the best answer to each question in four minutes. His choice is written on a separate record sheet. Discussion and self-scoring follow. The number of correct answers made by each pupil is converted into a G score by means of the table which follows each lesson and the G score is recorded on the record sheet and compared with the G grade and G age. Thus the tests are standardized.

3. *Practice Exercises in Reading*. Four Books: Main Thought, Details, Directions, Prediction. Bureau of Publications, Teachers College, Columbia University, New York.

The lessons are similar to the standard test lessons except that they do not yield a G score for each lesson, and that they do not mix several kinds of reading skills in each lesson but provide,

rather, a separate book for each skill. This permits the pupil to concentrate on a particular skill, if a previous diagnosis discloses a specific difficulty.

TEST LESSON 72

I have been asked to send a message from China to the children of the United States. I shall tell you a story I told to an American professor when we walked among the golden palaces of the Forbidden City in Peking, China.

Thousands of years ago, the King of our Flowery Kingdom received a gift of a wonderful pearl from the Emperor of India. While showing the precious gem to his nobles, it slipped from his fingers and rolled into a small, round, deep hole in the rock. Some tried to lift it out with long, slender strips of bamboo, but the gem fitted too snugly in the hole. When no one could think of a way to get the pearl, the king's joy over the gift turned to sorrow. Then a small lad, no older than you who are reading this, stepped forward and offered to get the gem. The king forgot his sorrow in laughing at a mere boy who thought he could do something the wisest man had not been able to do. What do you think the boy did? The next lesson will tell.

1. What was sent from China? (a) bamboo; (b) message; (c) pearl; (d) gift.
2. The story was told in (a) America; (b) India; (c) Peking; (d) a palace.
3. The hole was (a) shallow; (b) deep; (c) large; (d) square.
4. Who dropped the gift? (a) nobles; (b) emperor; (c) child; (d) king.
5. The lad caused (a) joy; (b) sorrow; (c) regret; (d) amusement.
6. The gem was sent to (a) the Flowery Kingdom; (b) the Forbidden City; (c) nobles; (d) an emperor.
7. What was golden? (a) king; (b) pearl; (c) emperor; (d) palaces.
8. What was sent from India? (a) pearl; (b) flowers; (c) message; (d) bamboo.
9. Which is the highest office? (a) king; (b) noble; (c) emperor; (d) professor.

No. right	0	1	2	3	4	5	6	7	8	9
G score	1.5	2.9	4.0	4.7	5.4	6.0	6.5	6.9	7.3	7.8

GRADE IV

1. *Pupil Activity Reader*, Book 5. All books are similar from this point on. The selection and test questions vary in length from one page to eight pages. Most sets of questions give practice on nine skills plus skills in reading varied types of material, newspapers, reference books, and the like, designated by an ap-

propriate symbol. These questions are all of the recall rather than the recognition type. The tests are not standardized.

2. *Standard Test Lessons in Reading*, Book 3.
3. *Practice Exercises in Reading*, Books 2.

GRADE V

1. *Pupil Activity Reader*, Book 6.
2. *Standard Test Lessons in Reading*, Book 4.
3. *Practice Exercises in Reading*, Books 3.

GRADE VI

1. *Pupil Activity Reader*, Book 7.
2. *Standard Test Lessons in Reading*, Book 5.
3. *Practice Exercises in Reading*, Books 4.

GRADE VII

1. *Experiments in Reading*, Book 1, Harcourt, Brace and Co., New York. These are standardized test lessons yielding a memory grade score for the remembrance of what has just been read, and a comprehension grade score for understanding of the same material which remains before the pupil for reëxamination. These two together direct the pupil whether he should read more rapidly or more slowly or at about his present speed.

2. *Let's Read*, Henry Holt and Company, New York. Diversified selections are followed by test questions to be answered and directions to be followed. The tests are not standardized.

3. *Flying the Printways*, D. C. Heath and Company, New York. This book contains elaborate developmental lessons on the psychology and physiology of reading, and diversified selections accompanied by directions and questions. The tests are not standardized.

4. *Reading for Understanding*, D. Appleton-Century Company, New York. Selections from texts for courses in the junior high school are followed by directions and unstandardized tests.

GRADE VIII

1. *Experiments in Reading*, Book 2.

GRADES IX TO XIII

1. *Experiments in Reading*, Book 3.
2. *Roads to Reading*, Harcourt, Brace and Co., New York. Unstandardized test lessons designed to be simple and especially attractive to slow classes in Grades VII, VIII, IX, or X.

3. *Study Type of Reading Exercise*, Bureau of Publications, Teachers College, Columbia University, New York. Twenty exercises that will give insight into the reading process, at the same time they provide practice in certain reading skills. The tests are not standardized.

4. *Reading for Skill*, Noble and Noble, New York. Unstandardized test lessons on reading for exact meaning, on rapid reading, on library and other skills related to reading.

GRADE XIII

You for College, Harcourt, Brace and Co., New York. Standard test lessons in reading based upon and included in a textbook in comprehensive guidance for freshman college students. This may be used with bright high school seniors who plan to enter college.

For a list of 225 graded devices for the remedial and classroom teaching of reading, the reader is referred to:

Russell, David H., Karp, Etta E., and Kelly, Edward I., *Remedial Reading Activities*, Bureau of Publications, Teachers College, Columbia University, New York.

Test Lesson Program in Mathematics.—As with reading, the sets of test lessons in arithmetic are numerous. A few of the more fully developed sets are described. Others appear in Chapter VII. Except in algebra, test lessons have not come into extensive use in the high school.

GRADES IV THROUGH VIII

1. *Standard Practice Tests in Arithmetic*, World Book Company, Yonkers-on-Hudson.

Since these tests cover the four fundamentals quite thoroughly, they are described as a sample of all such sets. A set of these practice tests consists of 48 stiff cards which make 48 lessons. Each lesson, except lessons 13, 30, 31, and 44 which are test cards, and lessons 45, 46, 47, and 48 which are study cards, contains just one type of example. The lessons begin with simple examples and gradually become more complex, each additional lesson representing just one additional difficulty. When the pupil has mastered the forty lessons, he has mastered all the difficulties in the addition, subtraction, multiplication, and division of whole numbers. There is one set of practice lessons for each pupil.

Along with the practice lessons comes a Student's Practice Pad for each pupil. The practice pad contains sheets of tissue paper. The pupil inserts a lesson card into the pad and under a sheet of tissue paper. This permits the pupil to see the example and at the same time do all work on the tissue paper, thus enabling the lesson card to be used from year to year. The student's practice pad also contains sheets upon which a pupil can keep a daily tabular and graphic record of achievement and progress.

Along with both practice lessons and practice pad comes a Teacher's Manual, which gives detailed instructions for the proper use of practice lessons and practice pads and warnings against their improper use by over-zealous teachers. The manual also gives much helpful advice about how to diagnose and remedy pupil defects in the four fundamental processes. The manual also contains record sheets which enable the teacher to keep a continuous record of each pupil's work.

The essential steps in the procedure of using these practice tests follow:

1. All pupils are given test card 13 which contains all the difficulties found in lessons 1 to 13. Each pupil slips the test card, examples up, under the topmost sheet of tissue paper in his practice pad. At the signal all begin work and continue until the signal is given to stop.

2. Pupils exchange papers and score each other as the teacher calls the correct answers.

3. All pupils who make satisfactory scores are excused from lessons 1 to 13. Sometimes the test is given twice to make results reliable. Sometimes the excused pupils may do something else until the backward pupils catch up or they may take the next test and the next until a point is reached where they need to study.

4. All pupils not excused from drill take lesson 1. If they make a satisfactory score on lesson 1, the next day they take lesson 2, and so on.

5. Those who fail on lesson 1 continue studying it and taking it until a satisfactory score is made.

6. As soon as a pupil finishes lesson 12 he takes test 13 again as final proof of his mastery of the preceding lessons. He may work on something else until the others catch up or he may proceed.

7. As soon as about 90 per cent of the class, including those who originally passed, have finished test 13, they take test 30. Those who pass test 30 are excused, and those who do not, drill upon lessons 14 to 30 as described above.

8. The teacher keeps a daily record of what each pupil achieves, watches to see that there is no cheating, makes diagnoses and applies remedies where they are needed and *only where they are needed*, stimulates good work on the part of all, sees that pupils keep their own records in good condition, and occasionally rescores the pupils' papers in order to keep their standard of scoring high.

All the regular lesson cards have answers on the back, hence pupils may score themselves or each other by simply turning the lesson card over and reinserting it under the tissue paper. The teacher's attention is thus freed for the real work of individual instruction, since no papers are handed in to her except those which the pupil himself judges to be perfect.

Practice tests individualize instruction. Mass instruction is highly inefficient, and this is particularly the case with skills. The interests of study, instruction, and supervision are identical. All focus upon study. Study is highly individual. Instruction must be equally individual if it is to be efficient. Mass instruction aims at everybody. It frequently hits nobody.

The amoeba has three types of reactions produced by three types of stimuli. There are, first, positive stimuli in the form of satisfying food and the like. The amoeba reacts by advancing toward these stimuli. The teacher uses positive stimuli to attract pupils toward good habits of work. There are, second, negative stimuli to which the amoeba reacts by retreating. The teacher uses negative stimuli to drive the pupil out of bad habits of work. There are, third, neutral stimuli which produce neutral reactions in the amoeba, for neutral stimuli do not stimulate at all and neutral reactions simply mean no reactions at all. It is the teacher's ambition to become so efficient that every word she speaks or move she makes will be a positive or negative stimulus depending upon her choice. But in mass instruction most of the stimuli are neutral stimuli. Our professor of literature was right when he said that teaching the class was "like trying to pour water from a gallon bucket into small-necked bottles." Most of his stimuli were neutral, partly because of lack of capacity on

our part, partly because he was employing stimuli which were neutral to most, negative to some, and positive to only a few. Individual differences are so great that wherever possible mass instruction should give way to individual instruction. Practice tests are a device for individualizing instruction. Without the aid of some such device individual instruction is impracticable.

Practice tests automatically adapt the work to the ability of each pupil and thus enable each pupil to begin at that point which means neither reteaching nor premature teaching. This is accomplished by means of the initial inventory tests. Test 13 serves this function in the case of the *Courtis Practice Tests*.

Practice tests permit each pupil to work according to his own methods and help him to find his best method. It is surprising how varied are the methods by which pupils learn such narrow functions as addition, subtraction, multiplication, and division. Kirby ¹ has shown not only that what is the best method for one pupil is not always the best method for another, but also that pupils frequently do not discover their best method and best rate of work until they are under the pressure of raising their score.

Finally, practice tests permit each pupil to advance at his own rate. Every study of the varied rates of progress for pupils in the same class has revealed the need of some teaching method which makes provision for individual differences in this respect.

Practice tests strengthen the purpose to improve. Practice tests motivate the learning process by making visible both distant and immediate goals and by providing a method whereby a pupil can measure his rate of progress toward these goals. Every pupil keeps a record of each day's achievement and draws a graph showing his progress. These provisions motivate through their appeal to basic instincts. The instinct of rivalry is so strong that work is turned into play by the simple process of introducing into it this element of rivalry. Practice tests not only make possible a rivalry between individuals, which is probably the world's most ubiquitous form of motivation, but they also make possible higher types of rivalry, namely, rivalry with one's own past record, and the rivalry of one group with another.

This provision of practice tests for the keeping of scores is

¹ Kirby, Thomas J., *Practice in the Case of School Children*, Teachers College, Columbia University, New York, 1913.

prerequisite both to intense effort and real happiness in school work. The games at which both children and adults work hardest and are happiest are invariably games where a score is kept. Generally speaking conventional education does not keep scores. A sort of score is occasionally reported, but these scores are purely relative. They do not show how much each individual has surpassed his previous record. They show which pupil is relatively best and so on to poorest. What stimulus is that to pupils who know they cannot hope to outstrip a more capable competitor? And what stimulus is it to the victor who knows that victory comes without much exertion due to his native superiority?

Practice tests motivate learning by throwing responsibility for promotion, or the attainment of the goal, upon the pupil. Every idle minute puts off the day when the goal will be reached and every industrious moment hastens the coming of the day, and what is important, the pupil is made to clearly perceive this intimate relation and is forced to recognize the fairness and justice of it. Just as certain as a pupil idles he will be punished and just as sure as he works he will be rewarded.

Practice tests secure a maximum of exercise. The second fundamental law of learning is, according to Thorndike, the law of exercise. When purpose is strong or when the law of effect is appropriately utilized and when exercise is abundant we have the optimum conditions for rapid progress. Here is the way I once taught addition to a class of forty pupils.

"Will each pupil copy on a sheet of paper the addition examples which I shall read to you, five examples to the row?" And then,

"Mary, you give orally the answers to the examples in the first row." And then,

"John, you take the second row," etc.

Each patiently or mischievously, according to his nature, waited until his turn came to begin. Only one pupil's neurons were exercising at a time, because I told each one just exactly where the preceding one stopped. Subsequent observations of other teachers have shown that my stupidity was not an isolated case. This one-out-of-forty sort of exercise is quite common. Had I used modern practice tests, probably without knowing it I would have multiplied my efficiency just forty times.

Practice tests facilitate aid and diagnosis. Practice tests bring swift aid to the pupil who needs it, and prevent teaching when it is not needed. Effort expended which brings no return in terms of progress brings discouragement. When discouragement reaches a certain stage effort ceases. Under ordinary conditions pupils sometimes remain for years undiscovered in the Slough of Despond. When the pupil's curve of progress ceases to rise to reward his effort, a teacher is needed. For the teacher to help at any other time would probably be to waste her time and injure the pupil. When to teach is instantly revealed by the curve of progress graphed by the pupil.

Practice tests facilitate diagnosis. Successful diagnosis requires the teacher to discover the exact location of the difficulty and the exact cause of the difficulty. Like tracer bullets, the pupil's daily scores leave behind a fiery trail which instantly reveals the location of the difficulty. The very following of this trail helps to eliminate probable explanations and thus facilitates diagnosis.

The chief danger from practice tests is not that they will cause too much emphasis upon drill, because the accompanying manuals allot a conservative time and constantly urge teachers not to exceed this time. The chief danger is that teachers will consider practice tests as something apart, so that the abilities developed by them will not function in life situations. The use of practice tests should grow out of genuine situations and should be continually associated with genuine situations. There comes a time in the execution of projects where the pupil realizes that his skill is inadequate. It is the function of practice tests to repair this inadequacy in the most economical and interesting way.

2. *Standard Test Lessons in Fractions*, Bureau of Publications. Teachers College, Columbia University, New York.

If we were to trace in rapid outline the evolution of educational measurement, we would need to show how invention after invention has emerged, been improved upon, and cumulated. Among these inventions are standard tests with standard instructions, norms, a convenient and accurate T scale for scientific purposes, a convenient and popular age and grade scale for general purposes, rapid survey tests, diagnostic tests, and practice tests where every lesson is a test and every test a standard test. The *Standard Test Lessons in Fractions*, with its accom-

panying *Rapid Survey Test in Fractions* has synthesized all of these discoveries into a complete inter-related program for the improvement of pupils' skill in common fractions. They, like the tests and test lessons in reading by Gates, exemplify the trend in the testing and teaching of the skills.

Thus the *Rapid Survey Tests in Fractions*¹ may be used both as diagnostic and survey tests at one and the same time. This test is so constructed that each of the three forms may be used as a survey test and all forms together as a diagnostic test. The series is provided with G, T, and age scores and norms, so that whether the purpose of the examiner is diagnosis or survey, all the techniques are provided.

When the *Rapid Survey Tests* show there is need for practice, the *Standard Test Lessons in Fractions* provide it. Every lesson of the thirty-six is a standard test with a G score for every possible number right. Every lesson is a diagnostic test in which the inspection of the pupil's mistakes will tell the teacher what his instructional needs are. Diagnosis in the series is continuous and is an integral part of the teaching and testing.

The *Standard Test Lessons in Fractions* provide not only for repetition of the same lesson when needed, but also for the steady advancement of those who attain the degree of success to be expected of them. Also every test lesson, besides drilling on the particular process for which it is especially arranged provides for continuous review as well, for at the beginning of each test lesson are four review examples, put there to keep the pupil continuously working on the processes given previously.

The creation of such test lessons should shortly make diagnostic testing unnecessary except in clinics. For to be as efficient as they should be diagnostic tests should provide a continuous diagnosis that is intimately related to the material being taught, and that is integrated with the teaching process. This is not practicable with separate diagnostic tests, which are necessarily a series distinct from the teaching material, for, if used for any purpose but testing, their value for diagnosis will be destroyed.

Table 1 lists other test lessons, practice tests, or instructional tests in mathematics, language, writing, and such skills.

¹ Bureau of Publications, Teachers College, Columbia University, New York.

CHAPTER XXI

DIAGNOSTIC MEASUREMENT

1. FUNCTION OF DIAGNOSIS

A principal or supervisor who assumes responsibility for a new school, or a teacher who takes charge of a new class is faced with the necessity of making two types of diagnoses: a general diagnosis of the initial condition and a more detailed diagnosis of the particular defects of classes or pupils.

One important function of the initial inventory is to prevent re-teaching of abilities which have already been taught. Ayres and others have estimated the tremendous financial cost to the public of the 33 per cent of retardation in the schools. Someone has computed that \$40,000,000 are spent annually re-teaching pupils. No one has been willing to estimate the loss to the retarded pupils.

Unfortunately the real retardation and its cost have been little studied. Retardation studies have called pupils retarded who were not retarded and overlooked retardation which was really present. This is still another fallacy which has resulted from a study of surface appearances and one more argument for the use of educational tests to increase visibility for the really significant factors. Most pupils who are chronologically retarded are not educationally retarded at all. The only true cases of retardation are pupils who are kept below the grade for which they are fitted by educational age or Gp. Most of the chronologically retarded are where they belong educationally or, to be more exact, they are usually a little accelerated. It is the chronologically accelerated who are usually most retarded educationally. Thus educational measurement justifies the rather queer conclusion that chronological retardation tends to mean educational acceleration. Contrary to usual thinking, the chief cost of re-teaching occurs with the latter rather than the former group of pupils. It is the chronologically accelerated who are educationally retarded and who are re-taught something they already know. The chronologically retarded are, on the whole, re-taught some-

thing which they failed to learn from one teaching. Of course, re-teaching these mentally inferior pupils is costly, but in the long run it is probably less expensive than to permit them to proceed without adequately mastering the prerequisites. The function, then, of the initial inventory is to prevent the cost to pupil and public of re-teaching what has really been learned.

A second function of the initial inventory is to avoid premature teaching. We have already seen how pupils are frequently started on a phase of the curriculum which, in the light of their measured capacity to learn, is too difficult for them. We saw again that pupils are frequently required to learn a portion of the curriculum before they have learned certain prerequisites in the hierarchy. The initial inventory will not only prevent such premature teaching in general, but will definitely point out for the guidance of both learner and teacher just where the pupil is most deficient and hence where he most needs help. Two of the great wastes in education are due to re-teaching or premature teaching. An adequate initial inventory will prevent both. Says Foote, "When pupils and teachers know where they are and where they are to go there is reason to believe that the journey will be accomplished; otherwise it is very doubtful." It is the function of the initial inventory to show pupils and teachers *where they are*.

The detailed diagnosis to discover the causes of defects is pictured by the tree Igdrasil. Carlyle describes Igdrasil as the ash-tree of existence which has its roots deep-down into the kingdom of Hela, whose trunk reaches heaven-high, and whose boughs spread over the whole universe, a tree which is the past, present, and future, and what was done, is doing, and will be done. A central ability or purpose in a pupil is a miniature Igdrasil. Its roots reach deep-down into the educational conditions of early days, and its boughs spread through all his mental life; it shows the past, the present, and the future, and what was done, is doing, and will be done.

One bough of reading ability reaches into reasoning problems in arithmetic. The initial inventory reveals that the ability to solve written problems is defective. It then becomes the business of diagnosis to locate the cause, and the cause of the cause, and the cause of the cause of the cause, and so on back to the teaching unit. In sum it becomes the task of diagnosis to trace a

miniature Igdrasil from leaf to root. In the illustration it is the task of diagnosis to discover that the cause of inability to solve problems is a defective reading ability, and that the cause of a defective reading ability is an inadequate vocabulary and so on. Thus the method of diagnosis is to trace abilities to their roots by means of standardized tests in order to discover just which ability or element of it exists out of standard proportions. This is the method of locating the underlying causes of defects.

The function of such diagnosis is to guide corrective measures. There is an inscription upon the monument which commemorates the arrival of the first white man at the Cumberland River in Central Tennessee. The inscription is to his wife and reads thus: "She shed a leading light along his path of destiny." Diagnosis is the veritable wife of remedial instruction. Without its guidance corrective instruction is absolutely "hit or miss," with but one chance to hit and several million chances to miss.

There are an enormous number of diagnoses being made in our schools daily. Some of these diagnostic measurements are vague and penumbral and some are quite exact. Every increase in the accuracy of the diagnostic measurements means an increase in the percentage of hits. To make these diagnoses accurate requires time, but so does teaching. Many teachers do not realize that a large per cent of their pupils have not advanced one iota as a result of a year's teaching in, say, fundamentals of arithmetic. Diagnosis would mean a net saving of time.

2. COMMON CAUSES OF DIFFICULTIES

Success as a diagnostician requires: (1) A knowledge of the usual causes of usual defects in the various abilities developed by the school. (2) Eyes to see and training or experience to interpret subtle behavior as evidence of the operation of known causes. (3) A technique which will bring otherwise invisible hints to the surface. (4) A knowledge of what remedial measures to prescribe for a given diagnosis.

Summarized below are certain basic causes which are responsible for many defects and whose operation is not confined to any one subject. Just as pestilences can usually be traced back to a few sources, so many diagnostic traits, irrespective of the abilities from which they start, lead back to a few basic causes, especially when the defect being diagnosed is an annoyingly per-

sistent one. Before anyone attempts diagnosis he should have a knowledge of the more common fundamental breeders of ability defects.

Insufficient Practice.—In some pupils a given ability does not function at all, simply because they have never studied to develop the ability, or it functions imperfectly because they have not had enough study and practice. This condition need cause no special concern for it is easily remedied. The time for real concern comes when a normal amount of study and practice fails to eliminate the absence or imperfection of functioning.

Improper Methods of Work.—There may be an optimum method of work. Pupils who differ in type or temperament may require different methods or again there may be an optimum method for all pupils. At any rate many pupils are working below par because they are employing ineffective methods.

A special case of improper methods of work occurs in those abilities where speed and quality are intimately related. It may be that a pupil's ability is functioning imperfectly because it is functioning either too speedily or not speedily enough.

Deficiency in Fundamental Skills.—Deficiency may mean either absence of sufficient skill or absence of sufficient transfer of skill to the new situation or both. The mental processes cannot flower into appreciation of literature, nor is the mind free to reflect upon the principles of history, geography, science, mathematical problems and other higher stages in education until the underlying skills are both made automatic and transferable. The youth who does not come from a cultured home and whose learning has been hastily grafted on an ignorant home training, is barely conscious of his own ideas when addressing a cultured audience and scarcely enjoys what he eats when dining with a cultured family. All his attention is concentrated upon watching lest he "gabble like a goose," or upon observing lest he use the wrong spoon. The pupil who stumbles in his reading halts in his history. The remedy is to make the basic skills automatic.

Absence of Interest.—The importance of interest or purpose in developing ability cannot easily be over-emphasized. There are more failures due to failure of interest than this world dreams of.

Physical Defects.—The diagnosis of any ability should carefully consider physical factors. In the case of many pupils food

for their minds will not facilitate their school progress nearly so much as food for their stomachs. No diagnosis should omit a careful examination of sense organs, particularly the eyes and ears. Just as "rivers of mercy do not flow into the world through rye-straws," so we do not have an educational flood when knowledge and experience must trickle through choked sense organs. Instruction cannot possibly be more than 50 per cent efficient when the child hears only 50 per cent of what is said to him and sees only 50 per cent of what he looks at.

Again, diagnosis should consider the condition of the pupil's response mechanism. What goes in through the sense organs must come out through the response organs before the educative cycle is complete. More improvement in molding, drawing, painting, writing, manual arts, and sports might conceivably be secured through correction of defects of muscular coördination than through direct instruction in the abilities in question.

Thy body at its best
How far can that project thy soul on its lone way.

Subnormal Intelligence.—Low native intelligence is the pre-eminent cause of ability defects. Intelligence is the very tap-root of Igdrasil. Just as injury to the tiny pituitary body causes stunted stature, marked adiposity, imperfect sexual development and other profound changes, so a defective intelligence casts its blight upon many or all abilities. Because of its ubiquity and its probable unimprovability, this cause of defects has special significance. Its importance is not always understood by the superficial diagnostician, because the superficial diagnostician does not carry the process of diagnosis far enough. Unsatisfactory work in history may be traced to imperfect reading ability. But why is the reading ability imperfect? In many instances it will be found that reading ability is imperfect because of low native intelligence. Whenever retardation is general, and whenever there is relative unimprovability, it is well to test for intelligence.

3. SPECIAL DIAGNOSTIC METHODS

Diagnostic Methods: Introspection by Pupil.—This method is so obvious and is so frequently employed that it needs neither

discussion nor illustration. Pupils frequently know not only the exact location of their difficulty but the cause of the difficulty as well. When the pupil is able to diagnose his own difficulty it is a waste of time and effort for the teacher to resort to the more elaborate methods yet to be described. Even when the pupil does not thoroughly understand his difficulty a conversation with him may give the more experienced teacher sufficient data to make a diagnosis.

Diagnostic Methods: Observation of Normal Work.—The commonest method of diagnosis is to get some hint from the behavior of the subject being diagnosed. When school was not in session we three brothers worked in the mines with our father. He was particularly expert in diagnosing the condition of the rock under which we worked and in detecting the imminence of danger. For this reason he was always assigned to the dangerous task of removing the last coal which supported the overhanging rock. As more and more of the coal was removed the weight of the million of tons of rock slowly settled upon the frail wooden timbers. They would become taut like the strings of a violin, so that flying splinters caused by the pressure made a sort of music. Occasionally a timber would break with a sharp sound like the crack of a rifle. Through it all father worked as though unhearing. Perhaps a week later he would say: "Get your tools, boys, and get out as fast as you can." We would go a short distance to a place of safety, lie down behind a car so as not to be struck by loose objects blown by the wind of the fall, and listen to the snapping of the props and the grinding of the mountain. As we grew older we, too, learned to interpret hints given by the rock. Here as with wild things in the woods it was diagnosis or death, and diagnosis from subtle behavior hints.

The teacher watches a pupil read who is having difficulty with reading. She observes that his eyes do not have three or four evenly-spaced brief fixations per line, but move forward, then jump back again, and act in a generally irregular fashion. Observation of this behavior aids the trained teacher to make a diagnosis of the difficulty. Another pupil is rarely able to complete an assignment in history. By observing his study the teacher notes that while reading his lesson he screws up his face, shakes his head, moves his lips, and tugs at his hair. This, too, is a hint to the perceiving teacher. Another pupil is very slow at

figures. The discerning teacher may construct a trial diagnosis by noting that he is counting with his fingers, toes, or tongue, and whispering as he adds: "Seven and six make thirteen, and thirteen and eight make twenty-one." Another pupil is having trouble with division of fractions. An examination of his written work may reveal that the source of his difficulty is failure to invert the divisor. Thus accurate, detailed, trained, and experienced observation of pupils in the process of normal work is one method of discovering the data upon which to base a diagnosis and prescribe corrective measures.

Courtis¹ has listed some arithmetical defects discovered by this diagnostic method. Along with the defects he gives an excellent statement of the underlying causes and suggests corrections:

1. Child's movements very slow and deliberate, but steady.
2. Child's movements rapid but variable. Adding accompanied by general restlessness, sighs, frowns, and other symptoms of nervous strain.
3. Child's progress up the column irregular; rapid advance at times with hesitation, or waits, at regular or irregular intervals. Often gives up and commences a column again.
4. Child stops to count on fingers, or by making dots with pencil, or to work out in its head the addition of certain figures.
5. Child adds each first column correctly, but misses often on second and third columns.
6. Child's time per example increases steadily or irregularly; particularly after two or three minutes' work; i.e., 15 seconds each for first five examples, 17 seconds each for the next five, 23 seconds for next two, 45 seconds for the next example, etc.
7. Child's habits apparently good and work steady, but answer wrong.

The diagnosis and correctives follow:

1. Slow movements may be due either to bad habits of work or to slow nerve action. In the latter case, the difficulty will prove very hard to control. It is almost certain that no amount of training will ever alter the nerve structure and so remedy the fundamental cause. But in all such cases much can be done to generate ideals of speed, to help the child to eliminate waste motions, and to hold himself up to his best rate.

¹ Courtis, S. A., *Teacher's Manual for the Standard Practice Tests*, World Book Co., Yonkers-on-Hudson, copyrighted 1915. Used by permission of the publishers.

In any case the procedure would be as follows: Ask the child to add the first example alone so that you may time him. Give him the signal when to start and let him signal when he has finished. Let him make several trials of the same example to make sure that he does not improve under practice. The teacher should then give the child the watch and let him *time the teacher in working the same example*. Comment on difference in child's and teacher's times. Then have the child write in small figures all the partial sums, as shown in the illustration.

— The teacher should again time the child, letting him *read to*
 30 15 himself the partial sums as rapidly as he can. This will, of
 46 course, give the minimum time in which the child could pos-
 26 9 sibly add the example. The time records of a child with true
 41 defective motor control will show slight improvement, if any,
 22 8 even with such aid, and probably the only procedure to follow
 97 in such cases is to lower the standard to correspond. Where
 13 there is a marked difference in time between the original and
 60 this last performance, the child will get, for the first time in
 7 its life, perhaps, a perfectly clear *conception* of what working
 61 at standard speed really means, as well as the *sensation* of really
 working at that speed. The teacher and child should then
 practice the same example over and over until the child can *without*
the crutches add it at the standard rate. Now the teacher can give him
 the whole test again, urging him to work at his best speed and compar-
 ing his results with the first result. The improvement made by ten
 minutes of this kind of work enables the teacher to say that a proper
 amount of similar study would produce the changes desired.

"But," some teacher will say, "will the child not learn the example by heart?" This is precisely what is desired. A perfect adder has learned so many examples "by heart" that it is impossible to make up any arrangement of figures that will be in any way new to him. The child in the same way needs to perfect his control over *each* example until he finally attains to mastery over *all*.

2. If the child gives evidence of nervous strain, check his speed, teach him to relax and to work easily and quietly. Get good habits of work first, then bring up speed and accuracy by degrees. The nervousness of a child is usually caused by social conditions, physical health, or temperamental bias. In any event it is difficult to control. Look out for a large fatigue factor in nervous children.

3. Irregular speed up the column may be due to either of two factors: lack of control of attention, or lack of knowledge of the combinations. The latter factor will be discussed in the following paragraph (4). Attention will be considered here.

There is a limit to the length of time that a person can carry on any mental activity continuously. As time goes on, the mind tends to respond more and more readily to *any* new mental stimulus than it does to the old. The mind "wanders" as it is said. The attention span for many children is six additions, for some only three or four, for others eight, or ten, and so on. That is, a child whose attention span

is limited to six figures may add rapidly, smoothly, and accurately, for the first five figures in the column, giving its attention wholly to the work. As the limit of its attention span is reached, however, it becomes increasingly difficult for it to concentrate its attention. The child suddenly becomes conscious of its own physical fatigue, of the sights and sounds around it. The mind balks at the next addition; it may be a simple combination, as adding 2 to the partial sum, 27, held in mind. It finally becomes imperative that the child momentarily interrupt its adding activity and attend to something else. If this is done for a small fraction of a second, the mind clears and the adding activity will go on smoothly for a second group of six figures, when the inattention must be repeated.

It should be evident that these periods of inattention are critical periods. If the sum to be held in mind is 27, there is great danger that it will be remembered as 17, 37, 26, or some other amount, as the attention returns to the work of adding. The child must, therefore, learn to "bridge" its attention spans successfully. It must learn to recognize the critical period when it occurs, consciously to divert its attention while giving its mind to remembering accurately the sum of the figures already added. This is probably best done by mechanically repeating to one's self mentally, "twenty-seven, twenty-seven, twenty-seven," or whatever the sum may be, during the whole interval of inattention. Little is known about the different methods of bridging the attention spans and it may well be that other methods would prove more effective. The use of the device suggested above, however, is common.

Giving up in the middle of a column and commencing again at the beginning is almost a certain symptom of lack of control of the attention. On the other hand, mere inaccuracy of addition (as 27 plus 2 equals 28) may be due to lack of control over the combinations. If the errors occur at more or less regular points in a column, and if, further, the combinations missed vary slightly when the column is re-added, the difficulty is pretty sure to be one of attention and not one of knowledge.

4. Hesitation in adding the next figure, when not due to attention, is usually due to lack of control of the fundamental combinations. In such cases, however, the hesitation or mistakes are usually repeated *at the same point* on subsequent additions. The teacher should understand that it "takes time to make mistakes," and whenever a lengthening of the time interval occurs, it is a symptom of a difficulty which must be found and remedied.

In this case the remedy is *not* a study of the separate combinations. It has been proved¹ that for most children time spent in study of the tables is waste effort; that the abilities generated are specific and do not transfer. A child may know 6 plus 9 perfectly, and yet not be able to add 9 to 26 in column addition except by counting on its fingers. *The combinations must be learned, of course, but they should be learned*

¹ See Bulletin No. 2, Department of Coöperative Research, Courtis Standard Tests, 82 Eliot St., Detroit, Mich. Price 15 cents. See also, *Journal of Educational Psychology*, September, 1914.

by practicing column addition. Follow the method outlined in paragraph (1) above, having the column added over and over again until both standard speed and absolute accuracy have been attained.

5. The sums of a child who is unable to remember the numbers to be carried, but whose work is otherwise perfect, will usually have the first column added correctly, as well as all single columns. Unfortunately, however, inability to carry correctly is usually a fault of children with weak memories for partial sums in the column. It is well, therefore, to test the carrying habits of any child that is inaccurate. Many children do not add the number carried until the end of the next column; it should, of course, be added to the first figure in the column. If necessary the number to be carried should be emphasized as by saying, when the sum of a column is 27, "carry 2" to one's self as the 7 is written. This is again a time-consuming device which should be adopted only as a last resort. The carrying should be an automatic, unconscious operation. Repeated practice on a few examples until the same become so perfectly familiar that a child's whole attention may be given to establishing correct habits of carrying will prove beneficial.

6. Marked increases in the times required for the successive examples of a test are an indication of a fatigue factor in the control of the attention. Some children are unable to carry on continuously a single activity, as adding, through even a four-minute time interval without a very great loss in power. Two courses are open to the teacher, one or the other of which is sometimes effective: one is to determine the exact length of the interval at which the child can work efficiently, and then try to extend the interval slightly each day; the other is to set the child at work on very long and very hard examples, and to lengthen the time intervals to fifteen or twenty minutes' continuous work. Difficulties of this type are hard to remedy.

Diagnostic Methods: Oral Tracing of Process.—There are difficulties the underlying causes of which would never come to light from an introspective inquiry on the part of the pupil or from mere observation of the pupil's normal work. The purpose of the diagnostic process is, of course, to induce the pupil to commit some overt act which will reveal the invisible causes of his visible defect. When neither his ordinary actions nor his written work offers a suggestion it is well to have the pupil go through the process orally. When I fail to make the class in educational measurement understand the computation of, say, the median, I find it advantageous to ask one of the students who is having trouble to come to the blackboard and compute a median orally for the class. The cause of the difficulty is thus quickly found.

Uhl used the oral-tracing method to discover the mental processes through which pupils go in adding and subtracting. The

old phrase: "Beat the devil around the stump" accurately describes how some pupils work. To quote Uhl:¹

The findings as to methods employed by pupils in "difficult" combinations is both interesting and significant. The following methods were found in the work of pupils who were tried out in the manner just described. A fourth-grade boy showed by slow work that the combination $9 - 7 - 5$ was difficult for him. When questioned, he showed that he used a common form of "breaking-up" the larger digits. In working the problem, he said to himself: " $9 + 2 + 2 + 2 + 1 = 16$ and 21 ." This shows that the $9 - 7$ combination was not known, but that the $16 - 5$ combination was, inasmuch as he arrived at " 21 " directly after having combined the other two numbers. Another boy of the same grade showed the same type of difficulty in a more pronounced form. He added 8, 6, and 0 as follows: "First take 4, then take 2, then add 8, and 4 makes 12, and 2 makes 14." In adding 9, 7, and 5 he said: "9 and 3 is 12 and 4 is 16 and 2 - 18; and 2 - 20; and 1 - 21." He broke into parts even so easy a problem as $3 + 4 + 9$, adding $9 + 3 + 2 + 2 = 16$.

A pupil from the fifth grade presented a quite different method of adding. In adding 4, 9, and 6 she explained: "Take the 6, then add 3 out of the 4. Then 9 and 9 are 18, and 1 are 19." Other problems were worked out similarly: one containing 3, 9, and 8 was solved as follows: "8 and 8 are 16 and 3 are 19 and 1 are 20"; 5, 6, and 9 as follows: "6, 7, 8, 9, and 9 are 18 and 2 are 20." This tendency to build up combinations of 8's or 9's continued in the case of another problem: 6, 5, and 8 were added thus: "6, 7, 8, and 8 are 16 and 3 are 19." Probably her first problem was worked similarly, but I had to have her dictate her method twice before I understood; she then gave it as quoted.

Methods which are quite as clumsy are found in the case of subtraction. One boy of the fifth grade was found to build up his subtrahend in the case of many problems. For example, in subtracting 8 from 37, he increased his subtrahend to 10, then obtained 27, and finally added 2 to 27 to compensate for the addition of 2 to 8. Likewise, in subtracting 7 from 30, he added 3 to 7 and proceeded as before. This boy knew certain combinations very well, but did problems containing other combinations by a method much harder than the correct one.

Even greater resourcefulness was shown by a fifth-grade boy who found the differences between some numbers by first dividing, then noting the remainder or lack of one, then multiplying, and finally adding to or taking from the result as necessary. For example, in subtracting 9 from 44, he proceeded as follows: "Nine goes into 44 five times and 1 less; 4 times 9 are 36, minus 1 equals 35." That is, this boy knew certain multiplication combinations better than he did certain subtraction processes; therefore, he used multiplication, making adjustments either upward or downward as demanded by the problem.

¹Uhl, W. L., "The Use of Standardized Materials in Arithmetic for Diagnosing Pupils' Methods of Work," *Elementary School Journal*, November, 1917.

Diagnostic Methods: Analysis of Responses to Survey Test Items.—There are many tests specially designed to facilitate diagnosis. But practically every standard test has some diagnostic value.

Using his *Reading Scale Alpha 2*, Thorndike made an unusually subtle analysis of pupil results to discover the causes for imperfect comprehension in reading. The following selected quotations¹ will increase anyone's respect for the mental process called *reading* and will show the problem a teacher faces who undertakes to teach or diagnose this complex ability.

It will be the aim of this article to show that reading is a very elaborate procedure, involving a weighing of each of many elements in a sentence, their organization in the proper relations one to another, the selection of certain of their connotations and the rejection of others, and the coöperation of many forces to determine final response. In fact we shall find that the act of answering simple questions about a simple paragraph . . . includes all the features characteristic of typical reasoning. . . .

In correct reading (1) each word produces a correct meaning, (2) each such element of meaning is given a correct weight in comparison with the others, and (3) the resulting ideas are examined and validated to make sure that they satisfy the meaning set or adjustment or purpose for whose sake the reading was done. Reading may be wrong or inadequate (1) because of wrong connections with the words singly, (2) because of over-potency or under-potency of elements, or (3) because of failure to treat the ideas produced by the reading as provisional, and so to inspect and welcome or reject them as they appear. . . .

In particular, the relational words, such as pronouns, conjunctions, and prepositions, have meanings of many degrees of exactitude. They also vary in different individuals in the amount of force they exert. A pupil may know exactly what *though* means, but he may treat a sentence containing it much as he would treat the same sentence with *and* or *or* or *if* in the place of the *though*.

The importance of the correct weighting of each element is less appreciated. It is very great, a very large percentage of the mistakes made being due to the over-potency of certain elements or the under-potency of others. . . .

To make a long story short, inspection of the mistakes shows that the potency of any word or word group in a question may be far above or far below its proper amount in relation to the rest of the question. The same holds for any word or word group in the paragraph. Understanding a paragraph implies keeping these respective weights in proper proportion from the start or varying their proportions until

¹ Thorndike, E. L., "Reading as Reasoning: A Study of Mistakes in Paragraph Reading," *Journal of Educational Psychology*, June, 1917.

they together evoke a response which satisfies the purpose of the reading.

Understanding a paragraph is like solving a problem in mathematics. It consists of selecting the right elements of the situation and putting them together in the right relations, and also with the right amount of weight or influence or force for each. The mind is assailed as it were by every word in the paragraph. It must select, repress, soften, emphasize, correlate, and organize, all under the influence of the right mental set or purpose or demand.

Consider the complexity of the task in even a very simple case such as answering question 6 on paragraph D, in the case of children of grades 6, 7, and 8 who well understand the question itself.

John had two brothers who were both tall. Their names were Will and Fred. John's sister, who was short, was named Mary. John liked Fred better than either of the others. All of these children except Will had red hair. He had brown hair.

6. Who had red hair?

The mind has to suppress a strong tendency for *Will had red hair* to act irrespective of the *except* which precedes it. It has to suppress a tendency for *all these children . . . had red hair* to act irrespective of the *except Will*. It has to suppress weaker tendencies for *John, Fred, Mary, John and Fred, Mary and Fred, Mary and Will, Mary, Fred and Will*, and every other combination that could be a "*who*," to act irrespective of the satisfying of the requirement "*had red hair according to the paragraph*." It has to suppress tendencies for John and Will or brown and red to exchange places in memory, for irrelevant ideas like *nobody* or *brothers* or *children* to arise. That it has to suppress them is shown by the failures to do so which occur. The *Will had red hair* in fact causes one-fifth of children in grades 6, 7, and 8 to answer wrongly,¹ and about two-fifths of children in grades 3, 4, and 5. Insufficient potency of *except Will*² makes about one child in twenty in grades 6, 7, and 8 answer wrongly with "*all the children*," "*all*," or "*Will, Fred, Mary, and John*."

After completing a thorough analysis of results from tests of pupil's ability to solve arithmetic problems, Monroe diagnosed many of the errors as due to inability to read the problems, inability to calculate accurately, and inability to reason correctly, which are in turn due to still more fundamental causes. According to Monroe,³ pupils' mental processes when reasoning incorrectly are fairly pictured by Adams' ³ description of how the

¹ Some of these errors are due to essential ignorance of "*except*," though that should not be common in pupils of grade 6 or higher.

² Monroe, Walter S., *Measuring the Results of Teaching*, pp. 154-72; Houghton Mifflin Co., New York, 1918.

³ Adams, John, *Exposition and Illustration in Teaching*, pp. 176-78.

canny Scottish pupils solved this freak of a problem: "If 7 and 2 make 10, what will 12 and 6 make?" The description follows:

A look of dismay passed over the seventy-odd faces as this apparently meaningless question was read. Everybody knew that 7 and 2 didn't make 10, so that was nonsense. But even if it had been sense, what was the use of it? For everybody knew that 12 and 6 make 18—nobody needed the help of 7 and 2 to find that out. Nobody knew exactly how to treat this strange problem.

Fat John Thomson, from the foot of the class, raised his hand, and when asked what he wanted, said:

"Please, sir, what rule is it?"

Mr. Leckie smiled as he answered:

"You must find out for yourself, John; what rule do you think it is, now?"

But John had nothing to say to such foolishness. "What's the use of giving a fellow a count ¹ and not telling him the rule?"—that's what John thought. But as it was a heinous sin in Standard VI (seventh grade) to have "nothing on your slate," John proceeded to put down various figures and dots, and then went on to divide and multiply them time about.

He first multiplied 7 by 2 and got 14. Then, dividing by 10, he got $1\frac{2}{5}$. But he didn't like the look of this. He hated fractions. Besides, he knew from bitter experience that whenever he had fractions in his answer he was wrong.

So he multiplied 14 by 10 this time, and got 140, which certainly looked much better, and caused less trouble.

He thought that 12 ought to come out of 140; they both looked nice, easy, good-natured numbers. But when he found that the answer was 11 and 8 over, he knew that he had not yet hit upon the right tack; for remainders are just as fatal in answers as fractions. At least, that was John's experience.

Accordingly, he rubbed out this false move into division, and fell back upon multiplication. When he had multiplied 140 by 12, he found the answer 1680, which seemed to him a fine, big, sensible sort of answer.

Then he began to wonder whether division was going to work this time. As he proceeded to divide by 6, his eyes gleamed with triumph.

"Six into 48, 8 an' nothin' over, — 2 — 8 — 0 an' no remainder. I've got it!"

Here poor John fell back in his seat, folded his arms, and waited patiently till his less fortunate fellows had finished.

James ² knew from the "if" at the beginning of the question that it must be proportion; and since there were five terms, it must be compound proportion. That was plain enough, so he started, following his rule:

¹ Scottish: Any kind of arithmetical exercise in school.

² The clever boy of the class.

“If 7 gives 10, what will 2 give?—less.”

Then he put down

$$7 : 2 :: 10 :$$

“Then if 12 gives 10, what will 6 give?—again less.” So he put down this time

$$12 : 6$$

Then he went on loyally to follow his rule: multiplied all the second and third terms together, and duly divided by the product of the first two terms. This gave the very unpromising answer $1 \frac{3}{7}$.

He did not at all see how 12 and 6 could make $1 \frac{3}{7}$. But that wasn't his lookout. Let the rule see to that.

Diagnostic Methods: Developmental History.—Developmental history is as useful a method for educational diagnosis as for medical or mechanical or any other form of diagnosis. Go to a doctor with an obscure physical defect and he will enquire about your total past. An automobile repairman asks you to relate just what you did to the car to put it out of order. Take a mentally defective child to a psychologist and he will comb the child's history to see if something in that past may not suggest a diagnosis. The developmental history not only goes back to the prenatal environment of the child, but to the life of the parents and grandparents. Many fundamental educational defects are not of recent origin. They have been cumulative. They have remained unnoticed for years. Their roots reach far back into the past. A successful diagnosis requires that these roots be traced back to their origins.

Diagnostic Methods: Contrast of Opposites.—Frequently a teacher does not succeed at a diagnosis simply because she does not know what are the customary causes of defects in the ability in question. Suppose, for example, that a pupil is not making satisfactory progress because his method of work is inefficient. A teacher who does not know what methods are and are not efficient is not likely to succeed with this diagnosis.

A diagnostic method which will help inexperienced teachers is to contrast opposites. The contrast may be between the best and poorest of the class, of pupils in one grade with pupils in a lower or higher grade. This method is to observe the two or three most successful pupils at their work and immediately after to observe the two or three most unsuccessful pupils, or to have both groups trace the process orally, or to test both groups and

analyze the results, or to use any other of the diagnostic methods upon both groups at the same time. Diagnosis by contrasting opposites will throw in relief the differences between competent and incompetent pupils and will thus facilitate diagnosis.

Diagnostic Methods: Detailed Diagnosis by Diagnostic Tests.—Purely diagnostic tests of a very elaborate nature have been developed, especially in reading, arithmetic, and spelling. Since a complete list of such tests is given in Table 1, only a few are mentioned here by way of illustration.

Gates has prepared four reading tests¹ for the primary grades and four for the elementary grades which reveal whether a pupil's reading deficiency is due to inability to comprehend main thoughts, comprehend details, follow directions, or predict outcomes or some combination of these.

In case the diagnostic process needs to be carried further he has provided a battery of tests¹ which yield a much more detailed diagnosis, showing whether the pupil is deficient in ability to name letters, give letter sounds, pronounce words, refrain from reversals, and the like.

In addition to these pencil-and-paper diagnostic methods, there are now available machines for testing hearing and vision as well as a mechanical method of analyzing reading, namely, the Ophthalmograph,² which photographs the eye movements on moving picture film.

Recently an ambitious experiment in remedial reading was undertaken, under Gates' general supervision, in the New York City schools. Children needing remedial attention, mostly in Grades II through IV, were tested. The most common deficiencies apparent were:

1. Limited oral and reading vocabulary.
2. Inability to attack new words.
3. Word-by-word reading.
4. Poor comprehension of what was read.

Diagnosis was followed by word games, coöperative stories, easy, success-assured, individual, seat-work stories based on units of *Gates Primary Word List*, supplementary reading assignments, frequent progress records, continuous diagnosis, much review and sympathetic, confidence-inspiring tutoring. In gen-

¹ Bureau of Publications, Teachers College, Columbia University, New York.

² American Optical Company, Southbridge, Mass.

eral, some were taught by the Gates method, some by the Monroe method, a few by the Fernald method, and most by a combination of methods.

For arithmetic we have Monroe's *Diagnostic Tests in Arithmetic*¹ and the Buswell and John *Diagnostic Chart for Fundamental Processes in Arithmetic*.¹ In their excellent, accompanying manual of directions are listed and illustrated some 121 different types of errors in the four fundamental processes. Most elaborate of all are the *Compass Arithmetic Tests*,² which cover in a series of tests about every aspect of that subject.

It has always been easy for teachers to differentiate between good and poor spellers among their pupils. It has, however, been much less easy to discover the cause for any pupil's spelling difficulties, and because of this fact remedial work in this subject has often met with little success.

Russell³ shows that while certain physical handicaps, such as certain types of defects in hearing and vision, contribute to difficulty in learning to spell, it appears that most troubles with spelling are the results of failure to learn to use a definite and preparatory technique of studying words. The good speller is characterized by the ability to see wholes clearly, and especially to see in the word whole a relatively small number of parts, such as syllables or phonograms. When a good speller learns, he surveys the word, searching for these familiar parts, and, having found them, he reviews the word and studies each part very carefully. Beyond that point he may or may not close his eyes and try to see these parts in a series in his mind's eye and then write them on paper, part by part, checking the written product with the original. Many good spellers take a number of such steps in the beginning stages but drop out one or more, such as the step of visualization, as they become expert. The poor speller seems to lack the ability to see the words clearly and define usable units in it. For example, he may see the word, "compliment," only as a rather long and complex figure. When he tries to spell, he simply looks at the letters in a series, as c-o-m-p-l-i-m-e-n-t. The good speller gets a picture of the word as a whole, in which he notes its special peculiarities. Then he finds a

¹ Public School Publishing Company, Bloomington, Ill.

² J. B. Lippincott Company, Philadelphia, Pa.

³ Russell, *Characteristics of Good and Poor Spellers*, Bureau of Publications, Teachers College, Columbia University, New York, 1938.

ing. These subjects are all treated in detail and illustrated with case studies.

The Gates-Russell *Spelling Diagnosis Tests* constitute an instrument for diagnosing the spelling difficulties of individual pupils. The battery includes the following tests: Spelling Words Orally, Word Pronunciation, Giving Letters for Letter Sounds, Spelling One Syllable, Spelling Two Syllables, Word Reversals, Spelling Attack, Auditory Discrimination, and Visual, Auditory, Kinaesthetic, and Combined Study Attacks.

Case study techniques for diagnosing the causes of difficulties of a more obscure nature are treated in the following references:

Baker, Harry J. and Traphagen, Virginia, *Diagnosis and Treatment of Behavior Problem Children*, The Macmillan Company, New York, 1935.

Bingham, Walter V. and Moore, Bruce V., *How to Interview*, Revised Edition, Harper and Brothers, New York, 1934.

National Society for the Study of Education, *Thirty-Fourth Yearbook—Educational Diagnosis*, Public School Publishing Company, Bloomington, Ill.

Symonds, Percival M., *Diagnosing Personality and Conduct*, Century Company, New York, 1931.

Thomas, Dorothy S., *Some New Techniques for Studying Social Behavior*, Teachers College, Columbia University, New York, 1929.

Wells, F. L., *Mental Tests in Clinical Practice*, World Book Company, Yonkers-on-Hudson, 1927.

As a sample of a thorough treatment of one elementary school subject the reader is referred to National Society for the Study of Education, *The Teaching of Reading: A Second Report, Thirty-Sixth Yearbook, Part 1*, Public School Publishing Company, Bloomington, Ill., 1937.

For a special reference which emphasizes diagnosis in regular high school subjects, the reader is referred to:

Ruch, G. M. and Stoddard, G. D., *Tests and Measurements in High School Instruction*, World Book Company, Yonkers-on-Hudson, 1927.

Green, Harry A. and Jorgensen, Albert N., *The Use and Interpretation of High School Tests*, Longmans Green and Co., New York, 1936.

CHAPTER XXII

EFFICIENCY OF PUPILS, TEACHERS, PRINCIPALS, AND SUPERINTENDENT

Do the results from standard tests given to a class reveal the efficiency of the teacher of that class? They do and they do not. They do provided certain conditions obtain. These conditions are roughly obtainable by an experimental control of the situation. They do not, because the conditions necessary for a just evaluation of a teacher's efficiency rarely obtain in the ordinary uncontrolled testing situation.

The use of tests for the guidance and diagnosis of pupils is so much more vital than their use to evaluate teachers that the former value should not be lost through antagonizing teachers in order to obtain the latter. For some time to come, at least, tests had better be used to measure pupils and not teachers, except in so far as teachers measure their own efficiency or coöperate in its measurement. When tests have reached a state of development where their use will lead to a just evaluation, the really efficient teachers will themselves demand to be rated by means of tests in order to escape another method whose accuracy is such that educators tolerate it only because nothing else has been available.

Since, however, teachers and supervisors are both likely to demand in the near future that their work be evaluated in a more scientific and hence more impersonal manner, there is summarized below the fundamental assumptions underlying a scientific procedure for rating and promoting teachers and supervisors as well as the steps in the process of making such ratings.

1. The pupil is the center of gravity or sun of the educational system. Teachers are satellites of this sun and supervisors are moons of the satellites.
2. All the paraphernalia of education exist for just one purpose, to make desirable changes in pupils.
3. The worth of these paraphernalia can be measured in just

one way, by determining how many desirable changes they make in pupils.

The pupil is the Alpha and Omega of all educational effort, and the center of gravity of the educational universe. Everything that Midas touched turned to gold. Everything that touches a pupil shows whether it is gold. Teacher, supervisor, principal, superintendent, United States Commissioner of Education, materials, methods, normal schools, this book, educational tests, the educational philosopher who confines himself solely to a contemplation of the ultimate, all these show whether they are gold or dross by the efficiency they show in altering the synaptic connections of this pupil's neurones. If no one of the above produces any desirable change in the pupil they are educationally without worth. Educational measurement is distinctive in that it must show the educational efficiency of all things and then in the last great experiment show whether it too has or has not value. Thus measurement alone possesses the power of self-destruction. And its worth like the worth of all else depends upon the amount and value of the changes it can produce in this pupil.

4. Hence the only just basis for selecting and promoting teachers is the changes made in pupils.

5. Teachers are at present selected and promoted primarily on the basis of their attributes, such as intelligence, personality, physical appearance, voice, ability in penmanship and the like.

6. No one has demonstrated just what causal relationship, if any, exists between possession of these various attributes and desirable changes in pupils. The relation between possession of certain attributes and the degree of favor of a teacher in the inspector's eyes is more evident. Dr. Chassell in her Ph.D. thesis determined the correlation between certain features of Ph.D. students of education and later success. She found that the score made in Ph.D. matriculation examinations at Teachers College correlated with success about .50. The quality of their Ph.D. dissertations correlated about .50. The letters of recommendation written about these Ph.D.'s correlated about .30. Their handwriting correlated .20, and their photograph .10. The following showed substantially zero correlation with later success: physical defects, type of locality of birthplace, age of reaching a given academic status, study abroad, size of family,

church relationship, reading knowledge of languages, and travel abroad. This study is more valuable in the present connection for the technique it exemplifies than for its conclusions. The subjects were not typical teachers but Ph.D.'s. The criterion of success was not demonstrated changes in pupils but the opinion of judges.

7. Scientific measurement itself is fair only when we measure the amount of desirable *change* produced in pupils by a given teacher. The measurement of *change* requires both initial and final tests. A final measurement is not sufficient, for the final status of the pupils was produced not only by their last teacher but also by all teachers who have taught the pupils, as well as all other environmental influences in the child's life.

8. Scientific measurement is fair only when we measure amount of change produced in a *standard time*.

9. Scientific measurement is fair only when we measure the amount of change in *standard pupils*. The following fable from the *Rose Garden* of the Persian poet, Sa'di, the "nightingale of Shiraz," is still true:

A king handed over his son to a teacher and said, "This is my son; educate him as one of thine own sons." The preceptor spent some years in endeavoring to teach him without success, while his own sons were made perfect in learning and eloquence. The king took the preceptor to task, and said, "Thou hast acted contrary to thy agreement, and hast not been faithful to thy promise." He replied, "O King! education is the same, but capacities differ."

The expectancy formula is a device for converting pupils, no matter what their intelligence and background into standard pupils.

10. Scientific measurement is fair only when the measurement is complete. Absolute completeness would require a measurement of the amount of changes made in children's purposes as well as their abilities. Absolute completeness is of course impossible and is in fact not necessary; partly because a chance sampling of the changes made will be thorough enough, and partly because teachers' skill in making desirable changes in, say, reading is probably positively correlated with their skill in making desirable changes in, say, arithmetic.

A practicable procedure whereby a teacher can get a rough idea of her efficiency follows:

1. Apply the *McCall Intelligence Test* and the *Educational Background Questionnaire* at the beginning of the school term, and compute an initial expectancy grade score for each pupil by means of the formula already described. Repeat both tests at the end of the school terms when the same pupils are about to depart, and determine a final expectancy grade score. The difference is the amount of growth in achievement we have a right to expect each pupil to make.

2. On or about the same dates apply the *Comprehensive Achievement Test*. Determine the initial and final grade scores in achievement for each pupil. The difference between them is the amount of growth in achievement produced.

The humanity of Pestalozzi and the sympathy for childhood of the good people who have followed in his train could not abide the dry-as-dust drill to which children were subjected. The reaction away from the drill subjects by certain educators is more than an emotional one. It is in part due to a real change in their conceptions of what is most worth while in education. They desire, and rightly so, a greater emphasis upon those virtues which have to do with civic responsibilities and other relationships. Since most existing tests measure drill subjects there is a grave fear that the widespread use of tests will merely increase the emphasis upon what they conceive to be the relatively less valuable abilities.

The attitude of these educators is wholly honest but substantially unwise. There is grave danger that they will use their ingenuity not to devise ways of tying the skills to children's purposes in such a way that drill will be interesting, but to undermine our conception of the tremendous value of these skills. Tom dreamt that he and many other chimney sweeps were locked in black coffins and that there came an angel with a golden key and unlocked the coffins and set them all free. The golden keys with which teachers unlock the minds of children are the basic skills. They are more valuable even than Virgil's golden bough for they open the very gates of life. The skills are valueless in themselves, and at the same time they are the indispensable prerequisites of all that is valuable in education. Like the centaur's tunic they cannot be torn off without carrying away the flesh and blood of the wearer.

Take reading, for example. Carlyle was not far wrong when

he said that all any school can do is to teach us how to read. Carlyle tells how Odin was credited with the greatest invention man has ever made, namely, the invention of letters whereby man may mark down the unseen thoughts that are in him. He tells of the astonishment of Atahualpa the Peruvian king; how he made the Spanish soldier who was guarding him scratch *dios* on his thumb-nail, that he might try the next soldier with it and thus ascertain whether such a miracle were possible. Odin deserved his deification.

Instead of objective tests causing an over-emphasis upon the skills to the detriment of purposes, their use will insure to skills just that emphasis provided by the curriculum and will prove to be the salvation of the higher values. When intelligently used tests are merely instruments for realizing the curriculum. Like poison, steam engines, fire or any other potent force they require intelligent control. We do not trust fire to infants, and if there exist anywhere educators who do not subordinate tests to their curriculum, they are still in their professional infancy and should not be trusted to use tests.

Tests will be the salvation of the higher values because of a natural human tendency to stress the tangible and visible. Just as a child will not put forth intense effort when he can see no results, so a teacher is not likely to spend much effort trying to develop a trait improvement in which neither she nor the child's parents can see. When a month's improvement in handwriting or composition is invisible, a year's improvement in unselfishness, even though very important will scarcely tip the scales of consciousness. It is human nature to fix our faith to form, hence so long as the average of human nature remains what it is, we must not expect it to expend effort in producing invisible, unrewardable improvements so long as it is permitted to produce visible rewardable changes. The moon pulls on the earth as well as on the sea but the earth tides interest few. Visibility and rewardability control the amount and direction of effort. The skills have been over-emphasized in the past, and always will be until we have either the thus-far-and-no-farther of tests or an educational magnifying glass which will make visible what has before been invisible.

All this is not said in opposition to the ultimate goal sought by such educators, but to their method of attaining it. In fact the

Comprehensive Achievement Test is recommended just because it is the most comprehensive test battery now available.

3. For each pupil subtract the expected growth from the obtained growth and record the result as either plus, zero, or minus. This is each pupil's efficiency measure.

The following story of Henry shows the importance of such measurements:

Henry is a relatively dull boy but his father doesn't know it. The teacher doesn't know it. The teacher considers him lazy. Two years ago Henry was in a class with pupils of his own age. Owing to his low intelligence he was hopelessly outclassed and as a consequence was failed by his teacher. When the father received the report, he and Henry had a dramatic session in the woodshed.

Henry repeated the work of the grade with another class which happened to be younger and duller than usual. As a result of this fortunate combination in his competitors, Henry's father received at the end of the year a good report of Henry's work. Henry's teacher is happy because she thinks she succeeded so much better with him than did his former teacher. The former teacher is happy because she thinks it was her courage in failing him that paved the way for a moral reformation. Henry's father is happy because he considers he knew just exactly the right stimulus to use to motivate Henry's study. Henry is happy because he is not as unhappy as he was a year before.

This year Henry is fighting a losing battle in competition with those who are intellectually superior. He already sees that he is headed straight for another failure which does not worry him, and another thrashing which concerns him greatly. Thus every other year Henry will receive his inevitable thrashing until he is strong enough to physically rebel.

Henry is not one child but a million children in this land of justice. These million are yearly subjected to such injustice because reports sent to parents are misleading.

The intellectually superior children suffer as much as or more than dull children, but their suffering is of a different type. Most gifted children are working far below their optimum level of efficiency simply because no one suspects their real possibility. Since they lead their classes without difficulty and hence secure all the rewards there is no motive to exceed their present rate of progress.

How may a just report be made? The foregoing calculations yield the fairest measure of the extent to which a pupil has progressed in proportion to how much he was capable of progressing. It is one of the measures which should be sent to parents.

A measurement of the efficiency of pupils is likewise useful in conferences with parents. Consider for a moment how much more useful a principal could make himself if he possessed for every pupil in his school the information shown in Table 21. Presented with an array of such impartial facts, the parent who came to scoff would remain to pray. Parents who came earnestly seeking means to coöperate would not go away empty of fruitful suggestions.

Fortified with such information the principal would be equally useful in conferences with teachers and pupils. Given such a detailed knowledge of the conditions in the school and of the problems with which each teacher is contending, the principal could add to his teacher's respect for his superior power, a respect for his superior knowledge. As the situation now stands the average principal must honestly confess that the rank and file teachers know far more than he of the real condition of the school. Finally there would be innumerable instances where such information would enable him more intelligently to confer with pupils, to deal with discipline cases, and to supervise the instruction of individual children.

Also the common practice of setting up no definite visible objective at all could not be expected to produce other than the current indifference toward improvement. We would question the intelligence of any adult who seemed to be in a great hurry if he did not know where he was nor where he was going. Without the initial measurements already recommended, children, as Foote of Louisiana points out, practically do not know where they are and without more definite objectives they do not know in any thrilling way just where they are to go. It may be a tribute to children's intelligence that they are listless and uninterested.

The common practice of setting up definite objectives which are not objectives at all but impossible ideals for the class can only produce discouragement. Either because of delusions of grandeur concerning their own efficiency or because of an irrational confidence in their pupils non-technically-trained teachers

and supervisors almost invariably set up impossibly distant objectives. Recently a group of unusually progressive teachers decided to set up objectives in composition. After months of study sample compositions were selected to mark the passing point for each grade. When these specimens were measured on a standardized composition scale it was found that the specimen selected to indicate the passing point for the fifth grade was of a quality which twenty-five per cent of sophomore college students could not equal.

The common practice of setting up a definite objective which is reasonable for the class as a whole, but which is the same for all the pupils in the class is almost equally bad. It violates a fundamental psychological law that pupils differ and differ greatly in both their initial ability and their capacity to make progress.

4. Average the pupil efficiency scores, regarding signs, to get the teacher's efficiency. Ignore pupils not present for all tests. Since the size of this figure will vary with the interval between the initial and final tests, it should be converted to a standard time of ten months. Thus the figure should be doubled when the interval is five months or multiplied by one and one-fourth when it is eight months.

This procedure may be used not only for the rating of teachers but also for selecting new teachers who are given a half-year's or year's trial.

Crude as the proposed method of selection is, it is fairer than present methods. The superintendent doubtless soliloquizes like Caliban upon Setebos as the applicants march before him with a sample of painstaking penmanship in one hand and an antique photograph in the other.

Am strong myself compared to yonder crabs
That march now from the mountain to the sea;
Let twenty pass and stone the twenty-first,
Loving not, hating not, just choosing so.
Say, the first straggler that boasts purple spots
Shall join the file, one pincer twisted off;
Say, this bruised fellow shall receive a worm,
And two worms he whose nippers end in red;
As it likes me each time, I do: so He.

Far too often a superior official finds himself in the awkward position of a certain school inspector. The teacher whose work

he delighted most to inspect, since she was the loveliest damsel on his circuit, was unfortunately an atrocious teacher. To report the facts would bring about her immediate dismissal and a consequent marked reduction of his pleasure when he made the usual rounds. After a prolonged struggle between his duty and his delight he decided to tell the truth. He reported her as being *pretty fair*.

5. Average the efficiency scores of all the pupils under one principal or superintendent to get a measure of the efficiency of the principal or superintendent and of the school and school system respectively.

The foregoing procedure is required to eliminate or make reasonably accurate allowance for the influence of (a) permanency of pupil population, (b) intellectual caliber of pupils, and (c) educativeness of pupil's background.

In response to a question as to how to measure teachers' efficiency when the age scale is the one used in the school, the author wrote as follows:

(1) At the beginning of the term apply a comprehensive battery of educational tests and compute class educational age. (2) At the same time apply an intelligence test such as the *Multi-mental Scale* or any other that is not composed of an excessive amount of educational test material, and compute class mental age. (3) Subtract class mental age from class educational age and attach sign. (4) At the end of the term apply another comprehensive battery of educational tests, and compute final educational age. Both batteries should test both basic skills and other important outcomes. (5) Compute class I.Q. Regard this I.Q. as a per cent, and multiply by it the months between the two educational tests. Add the product to the initial class mental age to get final class mental age. (6) Subtract final class mental age from final class educational age and attach sign. (7) If the algebraic difference found in (3) is about the same as that found in (6) the teacher's efficiency is about average. In proportion as the latter difference gets algebraically larger, the teacher's efficiency is above average. So compute the gain or loss of (6) over (3) and attach the proper sign to get the efficiency score. (8) Collect and make a distribution of these efficiency scores by grades and schools. (9) Study and broadcast the methods used by the most efficient teachers.

To avoid inducing a wrong educational emphasis it is important that all kinds of educational outcomes be sampled. Such test batteries are now available, as are other measures of learning capacity besides intelligence tests. Above all, the program should be so handled as to have the hearty and unembarrassed coöperation of the teachers, even though this means that the final results must be collected anonymously.

Crabbs,¹ in what is perhaps one of the most significant Ph.D. theses in education, applied and evaluated these methods of measuring the efficiency of teaching and supervision. Her critical evaluation led her to the following conclusions, among others:

1. In the rural schools there is a general tendency for efficient teachers of penmanship to be inefficient teachers of reading, arithmetic, and spelling. This tendency was not found for the urban teachers.

2. In general, a teacher's ability to teach a narrow skill can be predicted more accurately from knowledge of his ability to teach a wider skill than from knowledge of his ability to teach another narrow skill.

3. In general, a teacher's ability to teach a wider skill can be predicted more accurately from knowledge of his ability to teach another like type of skill than from knowledge of his ability to teach a narrow skill.

4. A teacher's ability to teach a wider skill can be predicted from knowledge of his ability to teach a like type of skill more accurately than his ability to teach a narrow skill can be predicted from knowledge of his ability to teach another narrow skill.

5. The correlation between ability to teach reading and supervisory estimate of that ability is .27 for rural and -.36 for urban teachers.

6. The correlation between composite objective efficiency and supervisory estimate of ability to teach everything except character is .33 for rural and -.25 for urban teachers.

7. The correlation between composite objective efficiency and supervisory estimate of ability to build character is .32 for rural and -.12 for urban teachers.

8. The correlation between composite objective efficiency and supervisory estimate of ability to teach in general is .32 for rural and -.26 for urban teachers.

9. The rural teachers were tested with the unpublished Steele-Herring test of professional knowledge and the scores made were correlated with composite objective efficiency scores. The correlation is .046. This indicates that there is no correlation between how much teachers know about modern educational ideas and practices as measured by the Steele-Herring test and how well they teach.

¹ Crabbs, Lelah Mae, *Measuring Efficiency in Teaching and Supervision*, Teachers College, Columbia University, New York, 1925.

10. The correlation between professional knowledge and ranking by supervisors for general teaching ability is .407.

11. In general the largest amount of accomplishment ratio (AR) change is in reading and penmanship. The smallest is in spelling.

12. The teacher-efficiency formula penalizes the teacher with high I.Q. pupils, provided these pupils also have high mental ages.

13. Contrary to conventional opinion, the teacher of bright pupils is benefited unless these pupils have high mental ages.

14. But the I.Q. of pupils affects only very slightly the validity of the teacher-efficiency formula.

15. The initial accomplishment ratio of pupils does not affect the validity of the teacher-efficiency formula.

Since Crabbs' efficiency measures were rather unreliable, thus making some of her conclusions necessarily tentative, it might be well to list here, pending final determination, the conditions which are likely, if present in excess, to make a teacher's efficiency score too small and *vice versa*:

(1) Pupils whose educational background score is low, unless it is used in the efficiency formula.

(2) Pupils whose abilities are extraordinarily high in the achievement test, thus reducing their opportunity to show gain.

(3) Pupils whose ability reaches beyond the grade in which they are classified and the curriculum provided for them.

(4) Pupils for whom the curriculum is too difficult or too uninteresting.

(5) Pupils who are often absent.

(6) Teachers who are often unavoidably absent, unless the substitute is a little more efficient than the regular teacher.

(7) Pupils whose curriculum doesn't fit the achievement test used.

(8) Pupils whose initial efficiency scores are markedly above normal.

(9) Pupils who have acquired previously a dislike for school or ineffective habits of work.

(10) A large class.

(11) Pupils whose ability falls on that level of the achievement test where the score intervals are wide because inaccurately scaled or because they are scaled in some units which are not *growth* units. Grade and age scores are growth units.

(12) Pupils whose growth in achievement is smaller than the true growth or whose growth in intelligence and background is

larger than the true growth, whether this be due to any of the foregoing causes, to unreliability of measurement, or to any other cause.

C. W. Odell in *A Critical Study of Measures of Achievement Relative to Capacity*¹ gives an excellent treatment of this much-discussed, important relationship, critically examining its validity and reliability and evaluating the numerous formulae which have been proposed for its measurement. Among the measures proposed are: Monroe's achievement quotient, Franzen's accomplishment ratio, Pintner's difference, McCall's F or T difference, McCall's G difference, Torgerson's efficiency quotient, Peter's accomplishment quotient, Otis' accomplishment quotient, Symond's index of effort, Nygard's accomplishment quotient, and Rand's sigma method.

The measurement of the efficiency of teaching can be more simply accomplished than by the use of capacity and achievement tests provided the person doing the measuring accepts the democratic-activity philosophy of education. All that is necessary is to apply the *School Practices Questionnaire* described in the preceding chapter. The supervisor's efficiency can be measured by applying this questionnaire twice and noting the increase in the score when expressed as increase per ten months. A zero increase would mean no efficiency as a supervisor of instruction unless it can be shown that his teachers would show a loss if he were not present.

Both these methods of measuring efficiency may, if desired, be supplemented by the subjective application of the criteria of good teaching presented in Chapter XVII, or by pupil rating.

Bryan² summarizes his study of pupil ratings thus:

The aims of study were as follows: (1) to determine how reliable and how valid are the pupil ratings of junior and senior high school teachers; (2) to determine how much agreement there is between the ratings of teachers by junior and senior high school pupils and administrators; and (3) to determine the effect upon ratings of such factors as (a) pupil mental ability as determined by standardized intelligence tests, (b) marks received by the rater from the teacher rated, and (c) sex of the pupils and teachers.

¹ Odell, C. W., *A Critical Study of Measures of Achievement Relative to Capacity*, Bureau of Educational Research, University of Illinois, Urbana, Ill.

² Bryan, Roy C., *Contributions to Education*, No. 708, Teachers College, Columbia University, New York.

The rating instrument used contains items relating to (1) teacher knowledge of subject, (2) discipline, (3) ability to explain clearly, (4) sympathy, (5) fairness in grading, (6) amount of work teacher does, (7) pupil liking for teacher, (8) amount pupils are learning, (9) work required of pupils, (10) pupil liking for subject, and (11) general teaching ability. Pupils had the choice of one of five descriptive phrases (superior, high average, average, low average, inferior) under each item and were asked to make favorable and unfavorable comments about each teacher.

Ratings were obtained from approximately 900 junior high school pupils and two junior high school administrators and from approximately 600 senior high school pupils and three senior high school administrators. The intelligence quotients and marks of the pupils of junior and senior high schools were obtained from the office records.

FINDINGS

The ratings of 40 junior high school pupils will produce reliability coefficients of .90 and above on all items except item 9. The ratings of 40 senior high school pupils will produce reliability coefficients above .90 on six out of eleven items. The number of ratings required to produce r .90 for the remaining five items are 48, 51, 59, 68, and 90.

Both junior and senior high school pupils are able to point out specific weak and strong points of a teacher's personality and methods to a degree that makes it worth while to obtain ratings on a series of items in addition to one rating on general teaching ability.

The points on which the ratings of pupils and administrators were compared revealed that: (a) the average ratings of groups of pupils are much more reliable than the ratings of a few administrators; (b) the amount of agreement between the ratings of senior high school pupil groups and administrators seems to exist in proportion to the degree of personal contact that the administrators had with the teachers and pupils; (c) on three items out of five, the average ratings of the junior high school principal and assistant principal agree more closely with the average ratings of pupils than the ratings of the principal agree with those of the assistant principal; and (d) administrators show more inclination than pupils to rate the same teacher about the same on all items.

No significant differences appeared between the ratings by pupils of high intelligence and those by the pupils lower on the scale of intelligence.

Both junior and senior high school pupils who received high marks showed a slight tendency to rate the teachers higher than the pupils who received the lower marks. There are many exceptions to this tendency, however, and in most cases the differences resulting from the higher ratings given by the pupils who received the higher marks are small.

Separate tabulation of the ratings by boys and the ratings by girls would appear to be worth while even though this procedure would

produce important diagnostic information relative to only a few teachers out of many.

The beta weights in the two regression coefficients indicate that the five items (out of ten) which have most positive relative weight in determining general teaching ability are items 1, 3, 6, 7, and 8 for senior high school teachers and items 1, 3, 4, 6, and 8 for junior high school teachers.

PRACTICAL VALUE OF PUPIL RATINGS

If ratings are obtained and used under the favorable circumstances indicated, the opportunities for creating favorable pupil reactions to teachers, subjects, and methods should be increased. The ratings would often reveal aspects of teacher behavior and methods to which pupils are not reacting favorably. These revelations could serve to emphasize the need for improvement and point out specific goals for improvement.

If Thorndike's "law of effect" means what it appears to mean, if Kilpatrick's theory of "concomitant learnings" is defensible, and if Briggs' teachings concerning emotionalized attitudes are true, all improvements in pupil reactions (that is, improvements in pupil liking and respect for teachers, subjects, and methods) would bring improved opportunities for accomplishment of both immediate and indirect objectives of teaching.

Pupil ratings in the hands of supervisors and administrators have the possible value of an additional means of evaluating the worth of teachers. The supervisor's rating form could consist of two divisions, one division containing items calling for the supervisor's estimate of teacher merit, and the other division containing probably quite different items calling for pupil estimates (averages) of teacher merit. The resulting estimate of a teacher's merit would thus be a composite of the ratings by both supervisor and pupils.

School surveys purport to determine the efficiency of schools. The comprehensive test program makes the typical survey in large part antiquated, for the usual survey does not go to the heart of the matter. It examines a multitude of factors which may or may not have anything to do with real efficiency, and it has been weakest where it should be strongest, namely in its evaluation of that for which everything else exists—the curriculum made actual to the children. If this is satisfactory, and the health and safety of the children are assured, and the cost is not excessive, and the staff is not being exploited to accomplish the foregoing, how the school board is selected, and how the staff is organized, and how the various functions are allocated, and how the teachers are chosen, and a thousand other such concerns of

most surveys are unimportant, for there may be a hundred patterns equally efficacious.

When surveying the total efficiency of a community, the grade scores on the *Comprehensive Achievement Test* should be compared with the grade score for the grade and, better still, the grade score for the age. When it is not desired to give the community credit for the inheritance it has provided for the pupils, the grade score in achievement should be compared with the grade score in intelligence. When surveying the efficiency of school board, superintendent, principals, supervisors, and teachers, the grade score in achievement should be compared with a combination of the grade score in intelligence and the grade score in background.

Those who wish to read further about efforts that have been made to test teaching efficiency are referred to the following reference:

Walker, Helen W., *The Measurement of Teaching Efficiency*, The Macmillan Company, New York, 1935.

BOOK SIX



SCHOOL MARKS AND REPORTS

CHAPTER XXIII

VARIETIES OF MARKING SYSTEMS¹

1. PERCENTAGE MARKING SYSTEM

Only a few of the numerous marking systems in use or that have been proposed will be described here. A common system is the percentage system, in which the student receives 100% if he answers all questions correctly, 90% if he answers 90 per cent of them correctly, 80% if he answers 80 per cent of them correctly, and so on down to that awful 70% or 75% below which no hope is left.

The superficial resemblance of these per cent marks from examination to examination has lured thousands of teachers into the belief that it makes the passing point on all examinations and all other points of equal or approximately equal significance, regardless of the great differences in the difficulty of tests and the ability of classes. Once the author saw the educational authorities of a state much upset because not enough teachers could be licensed to meet the needs of the state. Only a few teachers had made the required 75% in *one* of the subjects. The trouble lay not with the teachers but in the extra difficulty of that one examination—a difficulty that had not been and often cannot be foreseen even by the most experienced testers. Per cents should not be predetermined but should be attached, if they must be attached, after an inspection of the total scores made by all the pupils.

2. FREQUENCY DISTRIBUTION MARKING SYSTEM

The distribution marking system attaches per cents or letters only after an inspection of the frequency distribution of scores. Thus, the top 7 per cent of students in a given class receive a mark of A, or, say, 100% regardless of how many test items are answered correctly. The next 21 per cent receive a mark of B or 90%, the next 44 per cent receive a mark of C or 85%, the next

¹The author makes grateful acknowledgment of Dr. Harold H. Bixler's assistance in the preparation of Book Six.

21 per cent receive a mark of D or 80%, and the lowest 7 per cent are assigned a mark of F or 70%. These 7, 21, 44, 21, and 7 per cents make the steps between A, B, C, D, and F equal if the ability of the students in a class is distributed according to the normal frequency distribution (see Chapter XXXI). But the distribution may be far from normal. Furthermore, the system assumes, in effect, that all classes are of equal ability regardless of the grade or section—a violently unjustified assumption.

3. MEAN DEVIATION MARKING SYSTEM

Lindquist ¹ has described an ingenious modification of the distribution system which, as he recognizes, corrects one of its minor faults but none of its major ones. Abell, Sims, Ayer and Votaw, Mathews, and others have proposed similar marking systems. Lindquist undertakes to correct for the fact that in some classes or on some examinations the students are closely bunched whereas students in other classes or the same class on other examinations are widely dispersed. He suggests that the correction be accomplished thus: (1) by computing the average, i.e., mean, of the students' scores, (2) by getting the difference between the mean and every student's score, (3) by computing the mean deviation i.e., the average of these differences without regard to signs, (4) by adding twice the mean deviation to the mean to get the lower limit of the A group, (5) by adding two-thirds of the mean deviation to the mean to get the lower limit of the B group, (6) by subtracting two-thirds of the mean deviation from the mean to get the lower limit of the C group, and (7) by subtracting twice the mean deviation from the mean to get the lower limit of the D group. Scores below this limit receive E or F.

4. M SCORE MARKING SYSTEM

Russell ² accomplishes the correction of the same and other minor faults of the distribution system by means of the even more refined M score technique. He T scales (see Chapter XXXIII) the students' scores in a class and calls the resulting scaled T scores McCall or M scores. The M scores are kept and

¹ Hawkes, Herbert A. and Lindquist, E. F. and Others, *The Construction and Use of Achievement Examinations*, Houghton Mifflin Co., Boston, 1936.

² Russell, Charles, *Classroom Tests*, Ginn and Co., Boston, 1926.

used in that form instead of being translated into A, B, C, D, and F or per cent marks. To facilitate the computation and use of M scores, Russell has invented *The Classroom Scaler and Grader*¹ an ingenious combination of table and slide-rule features. But the M scale does not overcome the major defects of the distribution marking system.

5. G SCORE OR AGE SCORE MARKING SYSTEM

To overcome in large measure both the minor and major faults the author invented the grade score marking system or the age score marking system, if age scores are preferred to grade scores. Spence, Symonds, Somers, and Ellis have each proposed plans that are akin to this one in that they suggest making all marks comparable by means of an intelligence or other objective test.

About 1918, the author developed a method for securing comparability among scores on all group tests. By scaling total scores, these tests were made to yield mental age, reading age, spelling age, educational age, pedagogical age, promotion age, and their corresponding quotients. In the few years which followed, this age scale far outdistanced all others in popularity.

In 1922, he invented the grade scale especially for use in China. Its rapid and well-nigh universal adoption in the United States is a tribute to the extraordinary mental plasticity of educators and those engaged in intelligence and educational measurement. The United States and China are provided with more standardized tests than any other nations, and practically all of them now yield grade scores.

In 1923 he devised a method whereby informal classroom tests might also yield grade or age scores. The method was taught in his classes but withheld from publication because the original invention was too complicated for general use. The necessary simplifications to make the invention usable by all conscientious teachers in any nation provided with an appropriate calibrator test yielding grade scores or age scores have since been devised.

The essential steps in the normal operation of the grade score marking system from kindergarten through the university are:

1. Early in the term apply an intelligence test to secure for each student in the class a grade score in intelligence (Gi).

¹ Ginn and Co., Boston, 1931.

2. Arrange these Gi's in order of size, the highest at the top, and project them by increasing each by 0.1 for each month until the next promotion date.

3. Apply any teacher-made test or standard test in any subject, score it in any way, and arrange the papers in order of merit, the best at the top.

4. Assign the highest test paper the highest projected Gi. Assign the second highest paper the second highest projected Gi, and so on. Record these grade scores in the Teachers' Record Book under the subject covered by the test.

The remainder of Book Six is just an explanation, elaboration, illustration, and justification of this simple, four-step process, or two-step process after the first day, and of how all the grade scores thus made available can be used more effectively than the conventional marks they are designed to supplant for motivation, diagnosis, reports to parents, reclassification, promotion, transfer, graduation, and admission to higher schools,—in sum, for the guidance of pupil, teacher, principal, superintendent, parents, and personnel officers.

CHAPTER XXIV

CRITICAL EVALUATIONS OF MARKING SYSTEMS

In Table 25 are presented a list of criteria of a good marking system, together with application of these criteria in the evaluation of the percentage system, frequency distribution system, and the proposed grade score marking system.

For convenience, the criteria of a good marking system are set up in tabular form. In the three columns to the right, the existing systems and the proposed grade score system are rated in accordance with the scale of rating at the top of the table. These ratings represent the consensus of judgment of sixty students in a class in educational measurement at Teachers College, Columbia University, following a class discussion of proposed ratings.

Since there is much overlapping among the criteria, the reader may judge the three marking systems by all of the criteria or by any selection from them that appeals to him as being reasonable.

The percentage marking system satisfies fully only three of these criteria, partially only eight, receiving a total of 15 points. The distribution system scores 23 points, as compared with 64 points for the grade score system, which satisfies twenty-seven criteria.

There may be those who desire a fuller explanation and justification of these criteria. Take *Criterion number 2*, for example. It is perfectly reasonable to assume that a pupil's school career should not consist of a series of isolated years of work. There ought to be some way for him and his teacher to compare his achievement in one grade or year with his achievement in any other grade or year. If John's reading mark in the third grade is 3.2 and in the fourth 3.6, the amount of growth can readily be seen. Obviously the teacher also can see the amount of growth the pupil has accomplished.

Or take *Criterion number 11*, since intelligence tests have come into general use there has been a feeling on the part of parents

and teachers that a pupil's achievement should be judged in the light of his intelligence. Since marks, under the grade score system, are expressed in the same unit as intelligence test ratings, such judgment of a pupil's achievement can readily be made.

Criterion number 12.—Age-grade studies, which are now routine procedure in most systems, bring forcibly to the attention of the teacher the chronological ages of the pupils. A pupil's age or maturity affects his achievement; hence it is reasonable to expect the teacher to determine whether a pupil is achieving as much as should be expected, in view of his chronological age. The grade score system is the only one which makes satisfactory provision for this comparison.

Criterion number 13.—Reports to parents have been the subject of much discussion, especially among progressive educators, who, because traditional marking systems were unsatisfactory, have been recommending descriptive paragraphs instead. Public school administrators and teachers, who have to handle large classes without clerical assistance, feel that the individual letter or descriptive paragraph is not only impracticable, but that when marks are recorded in terms of grade scores, the teacher is able to give parents a more meaningful report, affording the basis for intelligent discussion of the pupil's problems.

Criterion number 14.—Despite many published articles and statistical studies, the average teacher still places great faith in the examinations that she prepares. Anything that enables a teacher to see how inaccurate and unreliable are her examinations is rendering her invaluable service. The fluctuations of a pupil's grade scores show up this inaccuracy plainly. Furthermore, Table 32 shows that there must be approximately one grade difference between an intelligence test score and a pupil's subject score based on a forty-minute examination, to be reasonably certain that the difference is a real one.

Criterion number 15.—Standard tests have become routine procedure in most progressive schools. Frequently there is no way to relate or compare these scores with a pupil's scores on informal examinations prepared by the teacher. Furthermore, since standard tests are usually both more valid and more reliable than informal tests, the administrator and teacher often desire to include these standard test results in the pupil's marks.

The proposed grade score marking system makes easy the comparison of standard test scores with the teacher's test results and enables the teacher to readily combine the two, since they are expressed in the same unit, i.e., grade scores.

Criterion number 18.—Parents frequently complain about differences in the system of classification of pupils into grades that exist within a given county or community. A pupil who is doing well in one fifth grade finds himself in difficulty when transferred to another fifth grade. It would seem that a reasonable criterion for any marking system would be that it should enable the administrator to standardize the system of classification throughout his district. Equal intervals of achievement between grades could then be maintained. Pupils transferring or entering from another system could be accurately classified, and the marks they bring with them could be used for this purpose.

Criterion number 22.—Although there has been some opposition to the sectioning or grouping of pupils into classes within a grade or subject, this is still a common practice and a convenient mode of adjusting the school to the needs of the pupil. The new marking system provides a wealth of data for sectioning pupils, even when no standard tests have been administered.

Criterion number 23.—Differences in teacher standards have long been a source of irritation to parents and administrators. In judging marking systems, a sensible criterion is the extent to which differences in teacher standards may be eliminated. The proposed grade score system is just and impersonal. The teacher does not have to decide whether a pupil should be given 80 or 90. Not only does this system protect the pupil from the injustice of too severe or too lenient marking, but it also largely frees the teacher from the strain of deciding upon marks.

Criterion number 25.—Unfortunately, some parents are more concerned about the marks their children receive than they are about the growth of the child. Since the grade score system is impersonal, parental attention is focused upon the growth of the child as shown by increase in his grade scores. Parents therefore are more often willing to let the school classify the pupil in that grade in which he can achieve the largest growth in score.

Criterion number 26.—Many educators believe that gradua-

tion or promotion to the next higher school should be based upon growth attained, rather than the number of years spent in school or the number of grades completed. The new plan makes this possible.

Criterion number 27.—One other criterion that appeals to the mental hygienist is the protection of the pupil from unfair pressure at home and in school. Since the teacher can make quick, easy comparisons of achievement with intelligence, she is not apt to expect too much from the pupil.

Criterion number 29.—A good marking system should provide such a system of comparable units as will enable the teacher or research worker to study scientifically many pressing educational problems. The grade score system utilizes a unit in common use for interpreting standard test results. The characteristics of the new plan as outlined in the discussion above give some indication of the possible comparisons and other uses in conducting researches.

Criterion number 30.—The grade score marking system provides a basis whereby colleges may evaluate, accurately for themselves and justly for the students, graduates from any high school in the United States or abroad. One of the chief reasons for Regents' examinations and college entrance examinations is the inability to make comparable the traditional marks brought from different high schools. Once the grade score marking system is in general use, these graduation and entrance examinations can be eliminated, thereby freeing the lower schools to serve their pupils instead of higher institutions.

Criterion number 33.—Graduation from grade to grade and especially from one school level to another has come to have much significance because it represents a form, albeit crude, of certification of attainment widely used by graduates in securing positions and by employers in evaluating applicants. This has placed a serious limitation upon the school, restricting its freedom to place pupils where they will grow the most. Perhaps certificates of attainment issued at periodic intervals, should replace graduation diplomas. They would certainly be more meaningful, would expedite the proper placement of graduates, and would restore to the school a much-needed liberty. Such certification can be based upon grade scores derived from examinations and the grade score marking system.

TABLE 25
RELATIVE MERITS OF THREE MARKING SYSTEMS

Key to Ratings

2—Reasonably Satisfactory

1—Partially Satisfactory

0—Unsatisfactory

CRITERIA	RATINGS		
	GRADE SCORE	PERCENT- AGE	DISTRI- BUTION
1. Enable pupil to compare his achievement with the average for all pupils in his grade and school	2	2	2
2. Enable pupil to see the amount of his growth in every subject from year to year	2	0	0
3. Enable pupil to compare his present achievement with his own past record for the year	2	1	2
4. Enable pupil to compare his achievement in different subjects and determine which need emphasis	2	1	2
5. Enable the teacher to compare the achievement of different pupils	2	2	2
6. Enable the teacher to see the amount of growth any pupil has made from year to year	2	0	0
7. Enable the teacher to compare the achievement of a pupil in different subjects	2	1	2
8. Enable the teacher or supervisory officer to compare the achievement of a pupil or class with the achievement of pupils or classes in the same grade in other schools or school systems	1	0	0
9. Enable the teacher to compare a pupil's record in one grade with his record in any other grade	2	1	1
10. Enable the teacher to section pupils within a class or grade	2	2	2
11. Enable the teacher to determine whether a pupil is achieving as much as you would expect, in view of his intelligence	2	1	1
12. Enable the teacher to determine whether a pupil is achieving as much as you would expect, in view of his chronological age	2	0	0
13. Enable the teacher to give a meaningful report to parents, so that teacher and parents may discuss intelligently the educational problems of the student	2	1	1

TABLE 25—*Continued*
 RELATIVE MERITS OF THREE MARKING SYSTEMS
Key to Ratings

2—Reasonably Satisfactory

1—Partially Satisfactory

0—Unsatisfactory

CRITERIA	RATINGS		
	GRADE SCORE	PERCENT- AGE	DISTRI- BUTION
14. Enable the teacher to see how inaccurate her examinations are.....	2	1	2
15. Enable the teacher or supervisory officer to compare a pupil's scores on informal examinations with his scores on a standard test.....	2	0	0
16. Enable the teacher or supervisory officer to combine scores on informal and standard tests....	2	0	0
17. Enable the teacher or supervisory officer to regulate the emphasis on different subjects.....	1	0	1
18. Enable the administrator to standardize the classification system throughout a community, county, state, or various colleges.....	2	0	0
19. Enable the administrator to transfer and accurately classify pupils from another school or school system.....	2	0	0
20. Enable the administrator to make such a classification of pupils as will secure and maintain equal intervals of achievement between grades.....	2	0	0
21. Enable the administrator to adjust the classification system to the intelligence of pupils in his school.....	2	0	0
22. Enable the administrator to section pupils into classes within a grade or within a subject.....	2	0	0
23. Enable the administrator to largely eliminate differences in standards of teachers, thus avoiding the injustice of too severe or too lenient marks..	2	0	1
24. Provide the school with a just and impersonal system, thus largely freeing the teachers from the strain of deciding upon marks and from the pressure of parents.....	2	0	1
25. Focus the attention of parents upon the growth of the pupil instead of upon a particular grade classification, thus enabling the school to classify			

TABLE 25—*Continued*

RELATIVE MERITS OF THREE MARKING SYSTEMS

Key to Ratings

2—Reasonably Satisfactory

1—Partially Satisfactory

0—Unsatisfactory

CRITERIA	RATINGS		
	GRADE SCORE	PERCENT- AGE	DISTRI- BUTION
pupil where growth in G score will be greatest.....	2	0	0
26. Permit school to base graduation upon growth attained instead of particular grade status, thus still further freeing the teachers from parental pressure.	2	0	0
27. Protect pupils from unfair pressure from home or school.....	2	0	0
28. Marks are always in numerical form, thus facilitating calculation.....	2	1	1
29. Enable the teacher or research worker to study scientifically many educational questions.....	2	0	1
30. Permit college and high schools to select pupils without entrance or Regents' examinations, thus materially freeing lower schools from the domination of upper schools.....	2	0	0
31. Permit schools consciously to set and intelligently administer minimum graduation and admission requirements..	2	1	1
32. Permit high schools and colleges to set admission requirements in terms of both achievement and brightness....	2	0	0
33. Permit exact and national certification of graduates in terms of achievement.	2	0	0
Total.....	64	15	23

CHAPTER XXV

PREPARING TO OPERATE THE GRADE SCORE MARKING SYSTEM

Step 1.—At the beginning of the school year, secure from some test publisher a non-verbal intelligence test for use in grades below the third, and the *Multi-Mental Scale*¹ *Form 1* for Grades III through XVI. This test is recommended because the divination feature embodied in it makes it one of the two single tests which may be used over a wide range of grades. The other—the *McCall Intelligence Test*² is even better, since it covers an equally wide range, is easier to score, and may be more readily used with pupils whose parents are foreign-born.

Step 2.—Administer the test to *all* pupils in the school. If any pupils are absent, they should be tested at a later time. Late entrants from outside the school system should be tested soon after they are admitted. The *Multi-Mental Scale* is so simply constructed as to be practically self-administering.

Step 3.—Score the papers, and, using the G tables provided by the publishers, convert the crude scores into G scores. Assume, for example, that the *Multi-Mental Scale, Form 1*, has been administered, and that pupil A has a crude score of 64 points. Turn to the G table, which is found in the *Manual of Directions* for this test. Find 64 in the "Score" column and read the G score opposite. It is 7.0, and is called a Gi or grade score in intelligence. It is interpreted to mean that the intelligence of pupil A is equivalent to that of the average pupil throughout the nation who is just beginning the seventh grade. A Gi of 7.1 is interpreted to mean that the pupil's intelligence is equivalent to that of the average pupil throughout the nation who has been in the seventh grade one month. G scores of 11.0 and 11.1 are interpreted similarly except that they refer to the eleventh grade.

Step 4.—Obtain a blank book or looseleaf notebook. If a Teachers Record Book is furnished by the school, any blank page

¹ Published by the Bureau of Publications, Teachers College, Columbia University, New York.

² Published by Laidlaw Brothers, 320 E. 21st Street, Chicago.

in it will be satisfactory. Prepare a Work Sheet similar to that shown in Table 26. At the top of the page, write the words "Work Sheet." On the first line record the following: Date of test, Names of pupils, Gi, and Projected Gi. In the column headed Gi record the Gi's of the class, in order of size, the largest first. In high school or college, several classes should be treated as a single class provided they are studying the same subject and are to take the same examinations. Skip a line between scores. In front of each Gi score record in the proper column the name of the pupil to whom it belongs and the date on which he was tested. (See Table 26, Columns 1, 2, and 3.) Note that if two or more pupils have the same Gi it makes no difference which is put first.

TABLE 26
SAMPLE WORK SHEET—ANNUAL PROMOTION

1 DATE OF TEST	2 NAME	3 Gi + 0.9	=	4 PROJECTED Gi
9-25-33	SI	6.5		7.4
9-25-33	SO	5.6		6.5
9-25-33	AC	5.5		6.4
11-13-33	MO	5.3		6.1 ^a
9-25-33	RO	4.8		5.7
9-25-33	AN	4.7		5.6
9-25-33	PR	4.6		5.5
9-25-33	MA	4.6		5.5
9-25-33	CA	4.6		5.5
9-25-33	CO	4.5		5.4
9-25-33	HA	4.5		5.4
9-25-33	ST	4.5		5.4
9-25-33	SW	4.4		5.3
9-25-33	CL	4.4		5.3
9-25-33	GR	4.3		5.2
10-7-33	HI	4.3		5.2
9-25-33	WI	4.2		5.1
9-25-33	WO	4.2		5.1
9-25-33	TH	4.0		4.9
9-25-33	LO	3.6		4.5
9-25-33	HO	2.6		3.5

^a Note that 0.8 is added to the Gi because pupil was tested November 13.

Step 5.—To the Gi (grade score in intelligence) add as many *tenths* as there are *months* until promotion time (see Table 27). The resulting scores represent the projected Gi, or the Gi as of promotion time, and should be recorded in Column 4 of the Work Sheet. For example, if pupils are promoted annually or are graduated from a course or subject annually, and if the test

TABLE 27

AMOUNTS TO BE ADDED TO G SCORES TO OBTAIN PROJECTED GI^a SCORES

DATE OF TESTING	ANNUAL PROMOTION	SEMI-ANNUAL PROMOTION
Aug. 16 to Sept. 15	1.0	0.5
Sept. 16 to Oct. 15	0.9	0.4
Oct. 16 to Nov. 15	0.8	0.3
Nov. 16 to Dec. 15	0.7	0.2
Dec. 16 to Jan. 15	0.6	0.1
Jan. 16 to Jan. 31	0.5	0.0
Feb. 1 to Feb. 15	0.5	0.5
Feb. 16 to Mar. 15	0.4	0.4
Mar. 16 to Apr. 15	0.3	0.3
Apr. 16 to May 15	0.2	0.2
May 16 to June 15	0.1	0.1
June 16 to June 30	0.0	0.0

^a The reasons for this and other procedures are given in a later chapter.

TABLE 28

SAMPLE WORK SHEET—SEMI-ANNUAL PROMOTION

1 DATE OF TEST	2 NAME	3 GI + 0.4	4 PROJECTED GI
9-25-29	SI	6.5	6.9
9-25-29	SO	5.6	6.0
9-25-29	AC	5.5	5.9
11-13-29	MO	5.3	5.6 ^a
9-25-29	RO	4.8	5.2
9-25-29	AN	4.7	5.1
9-25-29	PR	4.6	5.0
9-25-29	MA	4.6	5.0
9-25-29	CA	4.6	5.0
9-25-29	CO	4.5	4.9
9-25-29	HA	4.5	4.9
9-25-29	ST	4.5	4.9
9-25-29	SW	4.4	4.8
9-25-29	CL	4.4	4.8
9-25-29	GR	4.3	4.7
10- 7-29	HI	4.3	4.7
9-25-29	WI	4.2	4.6
9-25-29	WO	4.2	4.6
9-25-29	TH	4.0	4.4
9-25-29	LO	3.6	4.0
9-25-29	HO	2.6	3.0

^a Note that 0.3 is added to the GI because pupil was tested November 13.

were administered September 14, add 1.0 to the Gi's. If the test were administered near the beginning of the second month (September 16 to October 15), add 0.9 to the Gi's. For illustra-

tion of the procedure, see Table 26. For illustration of procedure when promotion is semi-annual, see Table 28.

If the school system operates on a time schedule different from the ten-month plan, use Table 27 anyway, regardless of the date of the beginning or ending of school. The error thus introduced will be negligible, except for schools that depart widely from the typical, such for example, as summer schools.

Step 6.—Prepare a place in the Teachers Record Book for marks, test records, etc. At the margin of the page on the left side of the book, record the names of the pupils alphabetically, last names first. Then rule this and the next six pages with vertical lines, about three-eighths of an inch apart. The pages on the right side of the book should be cut, as is customary in a printed Teachers Record Book, to avoid re-writing the names.

The first column should be headed Projected Gi (and date to which it is projected). In this column record the Projected Gi's.

TABLE 29

SAMPLE PAGE FROM THE TEACHERS RECORD BOOK

NAMES	Pro- jected Gi— JUNE	ARITHMETIC—GA							
		TIME 10 9-27	20 10-11	30 11-1	60 Qr. ^a	25 11-22	25 12-13	30 1-17	80 Qr. ^a
AC	6.4	5.7	7.4	5.6	6.2	5.6	6.1	6.5	6.1
AN	5.6	6.5	5.4	5.3	5.7	5.3	5.7	7.4	6.1
CA	5.5	5.3	5.1	5.1	5.2	5.1	5.1	5.3	5.2
CL	5.3	5.2	5.4	5.5	5.4	5.5	5.3	5.4	5.4
CO	5.4	5.6	5.5	5.5	5.5	5.5	5.5	6.1	5.7
GR	5.2	4.9	4.5	4.9	4.8	4.9	5.4	4.9	5.1
HA	5.4	5.4	3.5	5.2	4.7	5.1	4.9	5.2	5.1
HO	3.5	5.4	5.5	6.4	5.8	6.4	5.5	5.5	5.8
LO	4.5	3.5	5.3	ab	4.4	ab	ab	4.5	4.5
MA	5.5	5.1	5.7	3.5	4.8	3.5	3.5	5.4	4.1
PR	5.5	5.3	4.9	5.1	5.1	5.2	5.4	5.6	5.4
RO	5.7	7.4	5.4	5.4	6.1	5.4	5.6	5.5	5.5
SI	7.4	5.5	6.4	6.5	6.1	6.5	6.5	5.5	6.2
SO	6.5	5.4	6.5	5.3	5.7	5.3	6.4	5.7	5.8
ST	5.4	5.5	5.2	7.4	6.0	7.4	7.4	6.4	7.1
SW	5.3	4.5	5.6	5.4	5.2	5.4	5.1	5.1	5.2
TH	4.9	5.5	5.5	5.7	5.6	5.7	5.4	5.3	5.5
WI	5.1	6.4	5.3	5.2	5.6	5.2	5.2	5.4	5.3
WO	5.1	ab	5.1	5.5	5.3	5.4	5.2	5.1	5.2
HI	5.2	... ^b	5.2	5.4	5.3	6.1	5.5	5.2	5.6
MO	6.1	... ^b	5.5	5.3	3.5	4.8

^a Qr. Quarter This school issues report cards quarterly.

^b Pupil HI did not enter the school until October 7; pupil MO entered November 4.

Red ink may be used, if desired. Skip the next column. The remaining columns on this page should be assigned to the school subjects on which the teacher must mark.

At the top of the columns space should be left for recording the length of each test, in minutes, and the date.

For illustration of Step 6, see Table 29.

Since a high school and college teacher seldom teaches several subjects to the same identical group of students, one page will probably be sufficient for each class or each subject, if several groups are marked as a single class.

CHAPTER XXVI

MARKING EXAMINATION PAPERS IN TERMS OF GRADE SCORES

Step 7.—After an examination has been given the teacher should score the papers and arrange them in order of excellence, putting the best paper on top, second best next, and so on. If several pupils make exactly the same score, it makes no difference which one is given the advantage in position for they will get the same grade score in the end. In our sample class, an informal arithmetic examination was given on September 27. The pupils' papers were ranked by the teacher in the order shown in the first column of Table 30.

Step 8.—Turn to the Work Sheet (Table 26) and find the names of any pupils who may have been absent from the arithmetic examination. With a pencil draw a line, lightly, through (or put an x beside) the Projected Gi scores of absentees. These pencil lines must be erased as soon as marks have been assigned on this examination. If make-up tests are given before G scores are assigned to the other students this step is unnecessary. In our sample class, pupil WO was absent on account of prolonged illness, hence the teacher drew a line through Projected Gi score 5.1. This Gi is not used in any way until the absentee returns. Temporarily, it is ignored, as if this pupil had never been in the class.

Step 9.—Take the set of examination papers and record on the best paper, which is the first one on top, the first G score in column 4 of the Work Sheet (Table 26). On the second paper record the second highest G score. These scores now become the pupils' marks on this arithmetic examination. This mark is, for convenience, designated Ga, or grade score in arithmetic. The teacher must not think of these scores as the Gi scores of particular pupils. The logic underlying the procedure is this—the marks to be assigned a group of pupils should follow approximately the same distribution as the Gi scores, which represent the distribution of intelligence of the members of the group.

In our sample class pupil RO has the best paper on this arithmetic examination, hence he is assigned a mark of 7.4. Pupil AN, whose paper was second best, is assigned a mark of 6.5. See Table 30 for the marks of all members of the class. It is not necessary for the teacher to list the names and G scores as we have done in Table 30. As directed above, the marks should be placed on the pupils' papers.

Sometimes it happens that several pupils have the same number right, i.e., two pupils might have perfect papers. If an odd number of pupils (i.e., 3, 5, or 7 pupils) have the same score (number right), record the middle G score for all. If an even number of pupils (i.e., 2, 4, 6) have the same number right record the G score just above the middle.

This procedure is illustrated in Table 31. Pupil RO, having the largest number right (20) receives a mark of 7.4 (the highest Projected Gi). Pupils AN, WI, and AC have the same number right, i.e., 19. Obviously, it would not be fair to give one of them 6.5, another 6.4, and the third 5.7. Hence we take the *middle* score, which is 6.4, and assign it to all three. Similarly MA, GR, LO, and SW have the same number right, i.e., 9. Now there is no middle score, hence we arbitrarily and for con-

TABLE 30
ASSIGNMENT OF MARKS ON ARITHMETIC EXAMINATION

RANK ORDER OF PAPERS	RANK ORDER OF PROJECTED G SCORES
1. RO	7.4
2. AN	6.5
3. WI	6.4
4. AC	5.7
5. CO	5.6
6. SI	5.5
7. TH	5.5
8. ST	5.5
9. HA	5.4
10. HO	5.4
11. SO	5.4
12. PR	5.3
13. CA	5.3
14. CL	5.2
15. MA	5.1
16. GR	4.9
17. SW	4.5
18. LO	3.5
WO	absent

TABLE 31

ASSIGNMENT OF MARKS WHEN SEVERAL PUPILS HAVE THE SAME
NUMBER RIGHT

RANK ORDER OF PAPERS	NUMBER RIGHT	PROJECTED G _i SCORES	MARK
1. RO	20	7.4	7.4
2. AN	19	6.5	6.4
3. WI	19	6.4	6.4
4. AC	19	5.7	6.4
5. CO	18	5.6	5.6
6. HO	17	5.5	5.5
7. HA	16	5.5	5.5
8. TH	15	5.5	5.5
9. SI	15	5.4	5.5
10. ST	14	5.4	5.4
11. SO	13	5.4	5.4
12. CA	12	5.3	5.3
13. PR	11	5.3	5.3
14. CL	10	5.2	5.2
15. MA	9	5.1	4.9
16. GR	9	4.9	4.9
17. LO	9	4.5	4.9
18. SW	9	3.5	4.9
WO absent			

venience take the score just above the middle of the group, i.e., 4.9, and assign it to all four.

Step 10.—If a pupil is absent and the teacher gives make-up examinations but does not wish to delay assigning G scores to pupils in attendance until the make-up examination has been given, how shall G scores be assigned to the make-up examination? The following procedure is recommended:

- Until the make-up examination has been given, crude scored, and G scored, preserve a record of the crude scores made by the pupils in attendance together with the G scores assigned to these crude scores. This may be done by temporarily filing the examination papers on each of which appears its crude score and assigned G score or by making a list of the crude scores in one column and the assigned G scores in a column alongside. It is not necessary to write the pupils' names.
- Give a make-up examination and crude score it.
- Find on the list of preserved scores the crude score which is nearest in size to the crude score on the make-up examination. Give the corresponding G score to the make-up examination.
- Proceed similarly for other pupils if more than one make-up examination must be given.

If a pupil enters from without the school system, it will be necessary to administer the intelligence test at the earliest convenient date. Pupil MO entered November 4. On the test administered November 13, he scored a Gi of 5.3. By Table 27 we find that 0.8 must be added to 5.3 to obtain his Projected Gi, which is 6.1. This score (6.1) is, therefore, recorded in its place in the rank order (see Table 26). Note that the place of insertion will depend on the size of the Projected, not the original Gi.

CHAPTER XXVII

DIAGNOSTIC INTERPRETATION OF GRADE SCORE MARKS

Step 14.—After several examinations and ratings on themes and other activities have been recorded in the Teachers Record Book, the teacher should diagnose the work of the members of her class. Compare, for example, the scores of each pupil on the three arithmetic examinations given during the first quarter. For example, note the fluctuation in the case of pupil MA (see Table 29). Also, compare each student's scores in arithmetic with his Projected Gi. If a pupil's marks fall persistently and markedly below the Gi, the teacher should investigate the discrepancy between capacity (Gi) and achievement (Ga). To illustrate, possibly pupil MO's low scores are due to lack of effort and lack of interest after a late entrance. Pupil MA is not a consistent worker. Pupil SI lacks a good foundation in arithmetic. These are among the possible explanations. There may be two or more causes.

On the other hand, if a pupil's marks are consistently and markedly higher than his Gi, we should also study the situation. Pupil HO, for example, is probably a very hardworking pupil of less than average capacity. Such pupils deserve to be complimented, even though they may not rank high in the class.

Before making diagnoses, the teacher must be very careful to estimate the reliability of the examinations upon which the marks are based. Other things being equal, the longer and more objective the examination, the greater the reliability.

In view of the unreliability of examinations, the teacher must be careful not to attach significance to small fluctuations. To illustrate, fluctuations in the arithmetic marks of pupils CA, CL, CO, PR, SW, and WO are not significant (Table 29). The greater the fluctuation, the more significant it is.

When marks on several examinations are averaged, the results are more significant than marks on a single examination. For example, pupils AN, HA, and ST were doing better work in

arithmetic during the second quarter than during the first quarter (Table 29). In this instance we are comparing the marks for the Quarter, which are averages of marks on three examinations.

In the same way we must be cautious in comparing the Gi (grade score in intelligence) with any subject G score. In the case of pupil MA, it would not be safe to base conclusions upon the comparison of Gi with any one mark, such as that of September 27 or October 11. However, when we compare his Gi of 5.5 with the first quarter mark of 4.8 and with the second quarter mark of 4.1, we are fairly safe in saying that his achievement in arithmetic is not up to his capacity. In general, then, the more reliable the scores, the more significant are differences. Also, the greater the difference, the surer we can be that the direction of difference is significant. These big differences are the ones that need attention. Such problem cases should be referred to the school psychologist for special study and for more accurate determination of the Gi, since this also may be in error.

The question arises, "How big a difference is significant?" To answer this the teacher should first compute the total examination time represented by any subject G score. She should then refer to Table 32 and find in the first column the figure nearest to the time determined. Then read the figure opposite in the last column. This figure shows the number of months that the Gi—G subject difference must be in order to be significant. To illustrate, in Table 29, pupil HA has a Gi of 5.4 and a Ga for the first quarter of 4.7. The time represented is 60 minutes. By Table 32 we find that the difference, to be significant, must be .8 G or eight months. The actual difference is .7 G or seven months. We, therefore, cannot safely say that pupil HA is not working up to his capacity. Pupil MA has a Gi of 5.5 and a Ga for the second quarter of 4.1. The time represented is 80 minutes. By Table 32 we find that the difference, to be significant, must be .8 G or eight months. The actual difference is 1.4 G. Hence we are reasonably safe if we say that pupil MA is not working up to his capacity.

The first and last columns in Table 32 may also be used as very rough indications of when fluctuations in the same subject grade scores are to be regarded as accidental and when as significant. Thus pupil ST has a Ga of 6.0 for 60 minutes of examination in the first quarter and a Ga of 7.1 for 80 minutes of ex-

amination for the second quarter. Table 32 tells us that for an examination time of 70 minutes (average of 60 and 80) the difference must be .8 G to be significant. The difference is 1.1, i.e., 6.0 subtracted from 7.1. This is the only pupil showing a dependable difference from first to second quarter.

TABLE 32

SHOWING PROBABLE ERRORS AND SIGNIFICANT DIFFERENCES BETWEEN GI AND SUBJECT G SCORES FOR EXAMINATIONS OF VARIOUS LENGTHS

EXAMINATION TIME IN MINUTES	P E.	P E DIFFERENCE	SIGNIFICANT DIFFERENCE IN TERMS OF G
20	.45	.33	1.5
40	.3	.23	1.0
60	.26	.19	.8
80	.2	.176	.8
100	.2	.169	.7
120	.18	.165	.7
140	.17	.164	.7
160	.16	.164	.7
180	.15	.161	.7
200	.14	.160	.7

Step 15.—When standardized tests are given and when it is desired to combine scores yielded by them with G score marks, the scores on standard tests should be treated exactly the same as scores on any informal test. In the case of many standard tests it is possible to use tables provided by the authors and convert crude scores into grade scores. Such grade scores have approximately the same meaning and significance as the G scores based on teachers' examinations. However, in order to preserve the uniformity of the system, and comparability of scores from grade to grade and school to school, the crude scores should be converted to grade scores according to the plan outlined for informal examinations. This is neither necessary nor advisable unless the standard test scores are to be combined with the informal test scores.

Because of certain uses of standard test scores for comparison with norms and the like, the teacher may wish to set aside a page in the Teachers Record Book for comparing G scores obtained from standard tests with G scores obtained on her own examinations. One caution must be observed if standard tests are given during the term. Such scores are not comparable with

G scores based on informal examinations because the latter have been *projected* as of the end of the term. For example, pupil AC has a mark in arithmetic for the first quarter of 6.2. His Ga, according to the *Woody-McCall Mixed Fundamentals in Arithmetic Test*, administered on October 2, is 6.3. The two are not comparable, since the mark is *projected* to June. Also, the Ga on the standard test cannot be compared with his Projected Gi, because we should not compare his achievement in October with his intelligence in June.

Step 16.—Determine the G score age norm for each pupil by use of Columns 1 and 5 in Table 21. Find in Table 21 the pupil's chronological age, as of the date to which all G scores have been projected, and read the corresponding G score. This is his G score age norm, i.e., the G score we expect him to make in view of his age. This determination may be made at any time, but it is recommended for promotion time. Record the G score age norm for each pupil in the Teachers Record Book, and compare it with each pupil's grade scores in the various subjects. If his grade scores are lower than the G score age norm, we know that his achievement is not what we would expect in view of his age. If his grade scores are higher, we know that his achievement is greater than we would expect, in view of his age.

This computation is not an essential part of the marking system. It does, however, enable us to answer the question, "How well is this pupil achieving, in view of his age?" Grade scores on informal or standard tests, standing alone, do not answer this question. Two pupils may have the same grade score marks and, indeed, the same grade score in intelligence, yet one may be ten years old and the other twelve years old. The ten-year-old pupil is doing better, relative to his age, than the twelve-year-old. This should be recognized.

Another practical use for this type of diagnosis is in determining placement and promotion. Educators are giving much attention to pupils' progress through school, as indicated by age-grade studies and other analyses. Progressive educators are more and more coming to feel that the chronological and social age of the child must be considered in deciding where he shall be placed. Accordingly, if there is a question as to promotion, the G score age norm will indicate the grade where the pupil belongs, chronologically. To illustrate, pupil MA has low marks in

arithmetic; namely, 4.8 and 4.1 for the respective quarters. His marks in other subjects are also low. Shall he be promoted to the sixth grade? His chronological age is 13.2. By referring to Table 21 we read opposite 13.2 a G age score of 8.2. Obviously his grade scores in arithmetic of 4.8 and 4.1 indicate that he is not doing as well as we would expect in view of his age. His G_i is 5.5. Thus we have a pupil whose G score age norm of 8.2 argues for a seventh-grade classification, but whose intelligence and achievement are on fourth- and fifth-grade levels, respectively. With knowledge of these last two G scores, one will not make the mistake of placing him in a traditional seventh grade.

An analysis of the literature of the subject reveals that marks are being used to: (a) provide information to parents, (b) reward effort, (c) determine degree of mastery, (d) give credit for creative work especially, (e) provide incentive, (f) classify pupils, (g) give educational guidance, (h) predict future success, (i) give vocational guidance, (j) secure evidences of character and effort, (k) diagnose weaknesses, (l) determine promotions and admissions, (m) fix credits and honors, and (m) foster, by rating separately, citizenship, effort, health factors, and work habits.

The literature, referring mainly to conventional marks, helpfully warns that there is need to guard lest marks: (a) induce cheating, conceit, overwork, discouragement, and jealousy, (b) distract attention from real learning and become the pupil's major objective, (c) interfere with the free intercourse of pupils and teacher, (d) be based on something other than achievement or achievement in relation to capacity, (e) be unjust through great unreliability, (f) be a mixture of effort, conduct, interest, and personality with achievement.

For example, Sobel finds that pupils whose marks exceed their achievements as measured by standard tests are characterized by superior penmanship, attendance, punctuality, industry, perseverance, dependability, coöperation, ambition, as compared with those whose test scores exceed their marks, thus indicating that teachers allow such factors to unduly affect their marks on achievement instead of marking separately on such traits, and suggesting that teachers should check their marks periodically with standard tests.

CHAPTER XXVIII

RECORDS AND REPORTS TO PARENTS

Step 17.—At the time for reports to parents, average each pupil's marks in mathematics, and in the other school subjects. The report to parents should be in terms of grade scores. See sample elementary school report card, Table 33. The new sys-

TABLE 33
SAMPLE OF REPORT CARD TO PARENTS

Explanation of Marks

Marks are in terms of grade scores. These scores are based on the record to date, and represent the probable promotion time record. For example, a mark of 4.7 in reading is interpreted to mean that a pupil's achievement in reading is equivalent to that of the typical fourth-grade pupil who has been in the fourth grade seven months. Other marks are interpreted similarly. A fifth-grade pupil might have a mark of 4.7 in reading because in any grade there is always a wide range of achievement. Such a mark does not indicate need for demotion but it does show that he is below the average for the grade. On the other hand, if a fifth-grade pupil has a mark of 6.9 in history it does not necessarily follow that he belongs in the sixth grade. Such a mark does indicate that his work is above average.

Fifth-grade marks may be interpreted as follows, and similarly for grades above and below:

<i>Grade Score Mark</i>	<i>Interpretation</i>
7.0 and above	very superior
6.5 to 6.9	superior
5.5 to 6.4	satisfactory
5.0 to 5.4	fair
Below 5.0	unsatisfactory

GRADE 5	SEMESTER			
STUDIES	FIRST		SECOND	
	1st Qr.	2nd Qr.	1st Qr.	2nd Qr.
Reading	4.7	5.7	4.9	4.8
Oral Language	6.0	6.0	6.1	6.0
Written Language	5.9	6.0	5.8	5.7
Spelling	6.0	5.0	6.0	6.9
Penmanship	6.0	5.7	5.7	5.9
Arithmetic	6.2	6.1	5.3	6.0
History	6.9	6.7	4.9	6.0
Geography	5.7	4.9	5.0	6.0
Civics	5.9	6.3	5.1	5.8
Effort	5.8	6.1	6.0	6.3

tem should be fully explained to parents. They should be told either on the report card or elsewhere about the meaning of Projected Gi and its reliability. It is recommended that the Gi be not divulged to parents, except in very unusual problem cases and when the Gi has been quite reliably determined. The principal or the principal and teacher together will find it desirable to discuss the situation in a private conference with the parent. In most cases the parent will not question the general statement that the pupil is not achieving what he can. If necessary to reveal the basis of the statement, the Gi should not be referred to as an all-round measure of native intelligence, but rather as a specialized test of school capacity or scholastic aptitude. In our sample school, report cards are issued quarterly although only two quarters are shown in Table 29.

Step 18.—In addition to the grade scores, a descriptive statement and specific suggestions may be recorded on the report cards. For illustration see Cobb,¹ *The New Leaven*, or Reavis,² *Pupil Adjustment in Junior and Senior High Schools*. If desired, all of a pupil's subject grade scores, i.e., Ga, Gl, Gr, Gss, and so on may be combined into Ge or grade score in education. It is then possible to compute for each pupil Ge minus Gi. This is valuable for diagnosis and for conferences with parents.

Step 19.—Record the pupil's average marks on the Permanent Record Cards in the usual way. For illustration, see Table 34.

Of particular interest to high schools are the cumulative record forms developed by Wood and others under the Auspices of the American Council on Education. These, presented by Robertson in the *Educational Record*, January, 1933, may be adapted to permit the use of grade score marks.

In addition to the grade score marks provided for in Table 33, a space is suggested for a report on effort. Many administrators desire to acquaint parents with the pupil's achievement relative to his intelligence. This is particularly desirable when a pupil has low grade scores on account of a low Gi. For example, the

¹ Cobb, Stanwood, *The New Leaven*, pp. 220-23, John Day Company, New York.

² Reavis, William C., *Pupil Adjustment in Junior and Senior High Schools*, pp. 88-100, D. C. Heath and Company, New York, 1926.

norm for the end of the fifth grade would be 6.0, and yet pupil LO has a Ga of 4.5. We find, however, that his Gi is 4.5, so that he is really working up to his capacity. He deserves encouragement so effort should be marked "Satisfactory," or all the pupils may be ranked on the basis of the teacher's opinion, and G scores for Effort assigned in the usual way.

Only in the following rare cases should the effort mark be reported "Unsatisfactory":

1. Where the pupil's Ge is at least 1.0 below his Gi, and, in addition,
2. Where the pupil is known to be free from serious physical or health handicap, and, in addition,
3. Where the pupil is properly classified in view of his Ge, so that he really has a chance to grow. It is only when *all* of these conditions exist that we are justified in notifying parents that effort is unsatisfactory.

It is preferable, though not essential, to report all marks to parents in terms of grade scores. Among the reasons for doing so are the following:

1. Grade scores focus the attention of parents upon the growth of the pupil. Such marks are evidence of what he has accomplished, and do not have the appearance of mere opinion.
2. Grade scores impress parents as being just and impersonal.
3. Parents are able to see the amount of growth in every subject from year to year.
4. Parents may compare the achievement in different subjects, and determine which one needs more emphasis.
5. This plan is fairer to the slow pupil in that it enables the school to base promotion upon the growth accomplished rather than upon arbitrary status.
6. Pupils will be protected from unfair pressure, criticism, and nagging of parents.

We conclude, therefore, that the report should be in terms of grade scores. But, for the sake of those who prefer to retain the present system of reports, Table 35 shows how to convert G scores into letter marks. Appropriate per cents could be substituted for the letters.

TABLE 34
SAMPLE OF PUPIL'S CUMULATIVE RECORD CARD FOR ELEMENTARY SCHOOLS ¹

Name _____		
(Last)	(First)	(Middle)
(Other Identifying Data Not Reproduced)		

ATTENDANCE AND SCHOLARSHIP RECORD

Year	Semester	(Attendance, Name of Building, Name of Teacher, Not Reproduced)	Grade	Effort	Conduct	Reading	Arithmetic	Spelling	Language	History	Geography	Health	Pennmanship
1929-1930	1		5L	5.6	5.5	5.2	6.2	5.5	5.9	6.8	5.3	6.0	5.9
	2		5H	5.7	5.4	4.9	5.7	6.5	5.8	5.5	6.0	6.1	5.8

¹ From School Records and Reports, Report of the Committee on Uniform Records and Reports of the Department of Superintendence of the National Education Association, p. 250. Research Bulletin of the National Education Association, vol. V. No. 5, Washington, D.C., 1927. 350 pp.

TABLE 35

CONVERSION OF GRADE SCORE INTO LETTER MARKS

+ 1.0 or more above norm for end of grade	A
+ 0.5 to + 0.9 " " " " " "	B
- 0.5 to + 0.4 " " " " " "	C
- 1.0 to - 0.5 " " " " " "	D
- 1.0 or less " " " " " "	E
+ 1.0 or more " " " " " "	VS
- 1.0 to + 0.9 " " " " " "	S
- 1.0 or less " " " " " "	U

To illustrate, a fifth grade on the annual promotion plan would have an end-of-grade norm of 6.0. Table 35 then becomes Table 36 by making the necessary additions or subtractions, using 6.0 as a base. Thus 6.0 plus 1.0 equals 7.0, so scores of 7.0 and above are given a mark of A.

TABLE 36

SAMPLE CONVERSION TABLE FOR GRADE 5

7.0 and above	A	7.0 and above	VS
6.5 to 6.9	B	5.0 to 6.9	S
5.5 to 6.4	C	Below 5.0	U
5.0 to 5.4	D		
Below 5.0	E		

If pupils are sectioned according to ability, the conversion plan of Table 35 will tend to give low marks to all those in the slow section. Table 35 may be modified to suit the preferences of those using it.

Some time or other somebody in most schools sets out to discover how other schools make reports to parents. The teacher or principal can save himself much time by reading instead an article by Hill ¹ who analyzed 443 report cards from towns and cities of all sizes in practically every state and for grade levels from kindergarten through senior high school. His analysis showed:

1. That 45 per cent of the report cards were single- and double-faced cards, and 39 per cent were small folders.

2. That 80 per cent of the cards provided space for the teacher's signature, and 96 per cent provided space for the parent's signature.

¹ Hill, George E., "The Report Card in Present Practice," *Educational Trends*, February, 1934.

3. That 52 per cent of the cards were sent every two months, and 20 per cent every month.

4. That 80 per cent of the cards bore some sort of message from the school such as a request for coöperation, an invitation to confer with the teacher or visit the school, a request to sign and return the card, the importance of regular attendance, an explanation of marks, frequency of issuance, purpose of the card, aims of the school, bases of promotion, and explanation of failing marks.

5. That most primary cards did not list academic subjects, but provided instead for specific conduct habits, character traits, and health.

6. That a little over half of the high school cards did not list subjects but left these to be written in by hand.

7. That a five-point scale was the kind most commonly employed.

8. That more than 80 per cent of the cards above the primary indicated a failing mark. The per cent in the primary was 34.

9. That the mean number of character traits listed was 8.8 for the primary, 5.5 for the elementary, 5.1 for the junior high school, and 4.2 for the high school.

10. That the character traits tended to be specific in the lower grades and progressively more general.

11. That the most common traits were: *effort, conduct, coöperation, courtesy, obedience, persistence, reliability or responsibility, promptness, self-control, and attention.*

12. That, surprisingly, *health* was listed on less than half the cards, though the situation was better in the primary grades.

13. That certain cards contained unusual features such as, promotion certificates, comments by teachers, comments by parents, standard test scores, and extra-class activities.

14. That the trend is toward less frequent issuance of report cards, toward the use of informal letters with or without a formal card, toward the use of fewer steps in the marking scale, especially in the primary school, and toward the listing of specific rather than general traits. He makes three suggestions about character ratings that appear on report cards, namely, that the outcomes reported should be of prime importance, that they should be specifically defined in terms agreed upon by teachers and meaningful to parents, and that there should be positive as well as negative ratings lest the list look like an "inventory of delinquency."

15. That additional information will be found in Rowena Hansen's, *Report Cards for Kindergarten and Elementary Grades*, Leaflet No. 41, U. S. Department of Interior Office of Education, 1931.

16. That Hill promises to publish a set of recommended report reforms in the succeeding issue of *Educational Trends*.

Concerning the trend toward coarsening the scale, in the opinion of the writer, it is better to teach teachers, pupils, and

parents the important concept of probability and probable error than to abandon a convenient statistical mark for a cumbersome letter mark. Every pupil, teacher and patron needs to learn—and the sooner the better—that every measurement made in the world is not a point but a probable error blur—narrow with micrometer measurements and wide with mental measurements. If one twin gets 89% and the other 90%, the per cents don't need changing. It's the individual's way of looking at these two numbers. It is a blunder to fail to utilize such an excellent educational opportunity. Coarsening the scale does not eliminate the error. It makes it larger.

Some writers oppose any kind of report card. Here are some of the reasons that have been advanced for their complete elimination: (a) teachers are overworked with other matters, (b) some parents punish pupils for poor work, (c) many homes pay little attention to report cards, (d) the typical report card gives parents a distorted view of the school's objectives, (e) report cards serve as an incentive only to superior pupils. Hill reports that teachers and parents both favored overwhelmingly the substitution of teacher-parent conferences for *most* of the reports.

New developments in pupil report cards are given in *Educational Research Service*, Circular 4, 1934, National Education Association, Washington, D. C.

CHAPTER XXIX

CLASSIFICATION, PROMOTION, AND GRADUATION

In addition to their use for diagnosis, teachers depend upon marks largely to determine classification, sectioning, promotion, and graduation. In this chapter there will be presented a technique for using G score marks for these purposes. This technique has already been described in Book Three in connection with standard tests. It is sketched again here in order to show its application to records from informal tests, to clarify the procedure by multiplying illustrations and to avoid a fragmentary treatment of marks. The reader might well restudy Book Three at this point. Due to the proved unreliability of marks by subjects and to the desirability of permitting variation in students' subject emphasis, it is recommended that official "passing" or graduation be by grades or years rather than by subjects from kindergarten through the university. This does not preclude the possibility of repeating a particular subject or making other conventional adjustments. Even if graduation is by subjects solely, the techniques of this chapter may well be applied for the sake of the guidance records. It is vital that high schools and colleges use them toward the end of the final graduating year.

Step 20.—Prepare a page in the Teachers Record Book, to be designated the Summary Sheet. This page will have columns headed as follows: Name, Projected Gi, the name of the subjects which determine promotion, Gp, and Classification (see Table 37). When instruction is departmental this record will be prepared by the school rather than by the teacher, treating all freshmen, for example, as though they were a single class. If preferred, the Permanent Record Sheet may serve this purpose. As the time for promotion approaches, the teacher will transfer to this Summary Sheet the Projected Gi, and the term marks in reading, language, spelling, social science, and arithmetic, or marks in the subjects upon which the school bases promotion. These term marks are obtained by averaging the G scores for the

quarters, months, or marking periods. If the school marks quarterly and promotes annually, there will be four quarterly marks to average. See Table 29 for the first and second quarter marks.

Step 21.—Compute and Record G_p (grade score for promotion or placement). The formula for computing G_p is:

$$G_p = \frac{w G_i + w G_r + w G_l + w G_s + w G_{ss} + w G_a}{\text{sum of the } w\text{'s}}$$

where *w* signifies the weight to be used. The numerator may be extended to include any number of subjects.

TABLE 37
SUMMARY SHEET FOR GRADE 5

NAME	PROJECTED G _i	G _r	G _l	G _s	G _{ss}	G _a	G _p	CLASSIFI- CATION
AC	6.4	5.2	6.0	5.5	5.9	6.2	5.9	6
AN	5.6	5.0	5.2	5.0	5.1	5.9	5.3	6
CA	5.5	4.4	4.8	5.0	4.5	5.2	4.9	5
etc.	etc.	etc.	etc.	etc.	etc.	etc.	etc.	etc.

It is simpler and, on the whole, about as accurate to give each grade score a weight of 1. The formula then becomes

$$G_p = \frac{G_i + G_r + G_l + G_s + G_{ss} + G_a}{6}$$

To illustrate, in the case of pupil AC:

$$G_p = \frac{6.4 + 5.2 + 6.0 + 5.5 + 5.9 + 6.2}{6} = 5.9$$

Some schools mark on a large number of subjects. There will be some that are, relatively, less important for promotion than others as, for example, penmanship, music, nature study, drawing, and handwork. If it is desired to include all of these scores in the computation of G_p, the weights should probably be less than the weights for reading and arithmetic.

Another situation where different weights may be desirable is in a primary grade. Here reading is of major importance and should have, possibly, two or three times the weight given to number work.

To illustrate the computation, we will assume that in our

sample class the following weights have been assigned: ¹ intelligence 2, reading 2, language 1, spelling 1, social science 1, arithmetic 2. The formula then becomes

$$Gp = \frac{2 Gi + 2 Gr + Gl + Gs + Gss + 2 Ga}{9}$$

For pupil AC,

$$Gp = \frac{12.8 + 10.4 + 6.0 + 5.5 + 5.9 + 12.4}{9} = \frac{53.0}{9} = 5.9$$

Step 22.—Compute the mean (average) Gp for all pupils in the class and send it to the principal, together with a statement of the number of pupils enrolled.

Step 23.—The principal will then determine which classification table to use. The reader will observe that three classification tables are provided in Tables 40, 41, and 42. Table 40 may be used in a school which accomplishes 0.9 of a standard grade's work per year. In other words, such a school covers only minimum essentials. Table 41 is for use in a school which accomplishes one standard grade's work in one year. Most schools will use this table. Table 42 may be used in a school which accomplishes 1.1 of a standard grade's work in one year. This table may be used by schools in which the intelligence and achievement of the pupils are far above the average. Of course any one of the tables may be used by any one of the types of schools mentioned but to do so may involve drastic reorganization of the school. The procedure for determining the amount of achievement within a given school is described below. The amount of accomplishment is not necessarily proportional to the length of the school year.

The steps for determining which classification table to use are as follows: First record the mean Gp scores for all the pupils in each grade or year, together with the number of pupils in these grades or years. Then compute the total number of pupils in each grade and the total Gp scores; for example, in Table 38 there are two first grades. One has 38 pupils with a mean Gp of 1.7 or a total Gp of 64.6 (38×1.7). The other has 40 pupils with a mean Gp of 1.4 or a total Gp of 56 (40×1.4). Adding these we have a total of 78 pupils with a total Gp of 120.6, and dividing 78 into 1.6 we get an average of 1.5. Similarly, obtain

¹ For a fuller discussion of considerations influencing assignment of weights, see Chapter IX.

TABLE 38
CALCULATION OF DIFFERENCES BETWEEN GP'S AND NORMS

GRADE	1		2		3		4		5		6	
	No.	Gp	No.	Gp	No.	Gp	No.	Gp	No.	Gp	No.	Gp
No. on Roll and Mean Gp ..	38	1.7	40	2.9	40	3.8	35	5.1	40	6.2	35	7.0
No. on Roll and Mean Gp	40	1.4			29	3.5						
Totals	78	120.6			69	253.5						
Mean Gp		1.5		2.9		3.6		5.1		6.2		7.0
Norm Gp		2.0		3.0		4.0		5.0		6.0		7.0
Differences		-0.5		-0.1		-0.4		+0.1		+0.2		0.0
Total of Differences		-1.0 + .3 = -0.7										
Mean of Differences		-0.12										

the mean Gp for each grade. Then record underneath these the respective norms.

Then subtract algebraically, each norm Gp from the corresponding mean Gp, and record the differences with the proper signs. Total these differences algebraically and record the result opposite Total of Differences. (In Table 38, this figure is obtained by adding -0.5 , -0.1 , -0.4 , $+0.1$, $+0.2$, and 0 .) Compute the mean difference by dividing the total of differences by the number of different scores. In Table 38, -0.7 is divided by 6. The mean difference, -0.12 is interpreted thus: This school averages 0.1 Gp, or one month below the norm. Table 39 indicates which classification table to use:

TABLE 39
SELECTION OF CLASSIFICATION TABLE

IF THE MEAN DIFFERENCE IS	USE
Below -0.5	0.9 Classification Table
Between -0.5 and 0.5	1.0 Classification Table
Above 0.5	1.1 Classification Table

In our sample school, the mean difference (-0.1) is between -0.5 and 0.5 ; hence we shall use the standard 1.0 Classification Table. It happens that the principal uses or advises the teachers to use this table.

TABLE 40
0.9 CLASSIFICATION STANDARD TABLE

GRADE	CLASSIFICATION STANDARD	GRADE	CLASSIFICATION STANDARD
1L	1.5	9L	8.7
1H	1.9	9H	9.1
2L	2.4	10L	9.6
2H	2.8	10H	10.0
3L	3.3	11L	10.5
3H	3.7	11H	10.9
4L	4.2	12L	11.4
4H	4.6	12H	11.8
5L	5.1	13L	12.3
5H	5.5	13H	12.7
6L	6.0	14L	13.2
6H	6.4	14H	13.6
7L	6.9	15L	14.1
7H	7.3	15H	14.5
8L	7.8	16L	15.0
8H	8.2	16H	15.4

TABLE 41

1.0 CLASSIFICATION STANDARD TABLE

GRADE	CLASSIFICATION STANDARD	GRADE	CLASSIFICATION STANDARD
1L	1.5	9L	9.5
1H	2.0	9H	10.0
2L	2.5	10L	10.5
2H	3.0	10H	11.0
3L	3.5	11L	11.5
3H	4.0	11H	12.0
4L	4.5	12L	12.5
4H	5.0	12H	13.0
5L	5.5	13L	13.5
5H	6.0	13H	14.0
6L	6.5	14L	14.5
6H	7.0	14H	15.0
7L	7.5	15L	15.5
7H	8.0	15H	16.0
8L	8.5	16L	16.5
8H	9.0	16H	17.0

TABLE 42

1.1 CLASSIFICATION STANDARD TABLE

GRADE	CLASSIFICATION STANDARD	GRADE	CLASSIFICATION STANDARD
1L	1.6	9L	10.4
1H	2.1	9H	10.9
2L	2.7	10L	11.5
2H	3.2	10H	12.0
3L	3.8	11L	12.6
3H	4.3	11H	13.1
4L	4.9	12L	13.7
4H	5.4	12H	14.2
5L	6.0	13L	14.8
5H	6.5	13H	15.3
6L	7.1	14L	15.9
6H	7.6	14H	16.4
7L	8.2	15L	17.0
7H	8.7	15H	17.5
8L	9.3	16L	18.1
8H	9.8	16H	18.6

Step 24.—Determine and record each pupil's classification. To do this, look at the Classification Table selected (in this case Table 41), and recall the grade whose pupils are being classified (in this case 5H grade). Note the grade immediately above (in this case 6H), and find the corresponding classifi-

cation standard (in this case 7.0). Hold this classification standard in mind; return to the Summary Sheet; find all the pupils whose Gp scores are larger than this classification standard. Give these pupils a *double promotion*. Opposite their Gp scores record the symbol of the grade to which they are assigned. Even if the grade immediately above the grade being classified is missing, on account of pupils being transferred to another school, the same procedure will hold. However, pupils cannot readily be given a double promotion from a sixth grade to an eighth grade, when there is a junior high school.

In like manner note the classification standard for the grade below the grade to be classified. Turn to the Summary Sheet and find all the Gp scores which are *smaller* than this classification standard. Opposite these Gp scores record the symbol for the *present* grade. These pupils will not be promoted. To illustrate, when the fifth grade in our sample school is being classified, the classification standard for the grade below (4H) is 5.0. Any pupils with Gp scores *smaller* than 5.0 should be retained in the fifth grade, and 5 should be written in the Classification Column.

All other pupils will be promoted to the grade immediately above.

If promotion is semi-annual each half-grade is treated as if it were a grade. Thus, if the 5H grade were being classified, the grades above and below are 6L and 5L, respectively.

Finally, reconsider each pupil denied promotion and each one doubly promoted to see if there are any very special circumstances such as maturity, health, and parental attitude which forbid the proposed action. Also consider whether there are certain subjects which should be repeated, skipped, tutored or whether other educational adjustments are advisable.

After promotion has been decided, if the school is large, pupils may be classified into sections within each grade or year. A simple procedure follows: Take all the Gp's of all the pupils assigned to a particular grade and arrange them in order from highest to lowest. Determine the number of pupils to be assigned to each class or section. Count down the list of Gp's until that number of pupils has been counted. The appropriate section designation can then be marked in the Classification Column.

There are those who contend that pupils should be sectioned into classes within a grade in the same way they are classified into grades, namely by their total general status, i.e., Gp. Those who accept this contention should form sections within the grade or within the class according to the foregoing procedure. The author is inclined to recommend it because of its reasonableness and simplicity and because parents are accustomed to and are unlikely to object to the principle underlying it.

There are others who take the position, and quite reasonably too, that sectioning should be on the basis of brightness, i.e., *rate* of growth. Of course, since the highest Gp's are usually made by the youngest pupils in the grade, the highest section by the above procedure will include most of the bright, fast-moving pupils and conversely for the lowest section. But not *all* the brightest will get into the highest section and not *all* the slowest will get into the lowest section. Those who prefer to section by rate of growth may use the following procedure: (1) Convert the Gp of each pupil in a given grade into an age score expressed in months by the use of Table 21, Columns 1 and 5.

The Gp is found in Column 5 and the corresponding age score read in Column 1. (2) Divide the age score by the pupil's chronological age in months at promotion time and multiply the quotient by 100. This gives a Promotion Quotient. (3) Arrange Promotion Quotients in order of size and count off the pupils to be placed into sections. Or the same result may be secured more simply by sectioning on the basis of the difference which results when G age (see Chapter X) is subtracted from Gp and signs are retained.

Since there are excellent arguments for the two procedures described in the two preceding paragraphs, perhaps an average of both is better than either—an average that gives equal weight to general status and rate of growth in determining sectioning. This procedure follows: Arrange the names of the pupils in each grade in order of chronological age, beginning with the youngest. Then, on another sheet, arrange the Gp scores in order of size. Assign the highest Gp to the youngest pupil, and so on. These scores are now known as G age scores. Average each pupil's G age score with his Gp and call the average the grade score for sectioning. The group may then be divided into the necessary number of sections.

Step 25.—Graduate, certificate, and guide onward those whose grade scores warrant it. Pupils may be graduated from a six-year elementary school just as though they were being promoted from Grade VI to Grade VII, from an eight-year elementary school as though from VIII to IX, from a junior high school as though from IX to X, from a senior high school as though from XII to XIII, and so on. Of the schools which adopt this general plan, some will elect to certificate graduates on the basis of Table 40, others Table 41, and others Table 42, depending upon or regardless of the general intellectual level of the school. Unless the next higher schools provide for different levels of ability it is best to graduate by that one of the three tables which best fits all the schools which feed into the next higher schools. In a wealthy residential, or a predominantly Jewish, or a professional community Table 42 is likely to be indicated by the Gp's. In a foreign-born factory community Table 40 will probably be indicated by the Gp's. Typical communities will require Table 41.

If, as so often happens, the next higher school is adapted only to the best pupils from the lower schools and educational authorities elect to maintain these standards, then graduation must be by Table 42 or else graduation may be by Table 40 or 41 and guidance into the higher school by Table 42 or an even stricter table which may easily be constructed.

Thus, public or private high schools or colleges could stipulate the minimum Gp for admission, and so dispense with entrance examinations. These Gp's are reasonably comparable for all schools in the nation and are, possibly, more valid than entrance examinations. Since, within certain limits, the Gp, like the score on entrance examinations may be raised by repeating grades and thus adding the benefit of both extra study and maturing, the high schools or colleges which desire students who are able and also bright may stipulate both a minimum Gp and a minimum Promotion Quotient or, what is equivalent, a minimum excess of Gp over G age. The last two give the relation of age and achievement and hence either is a good index of brightness. A Promotion Quotient of 100, or a Gp minus G age of 0, indicates typical or average brightness.

A pupil promotion or graduation Gp may be his Gp for the last semester, last year, last two years, last three years or more.

If some of his preceding Gp's are averaged with his final Gp, they should first be projected to the date of the final Gp by adding 0.5 for each semester or 1.0 for each year intervening between the final Gp and the Gp being projected, or, more exactly, this 0.5 or 1.0 can be increased or decreased, as previously described, in proportion to the pupil's Promotion Quotient. How far back should we include Gp's? Should Gp's for years when the pupil was ill or often absent be omitted, or even the last year under some special conditions? The author prefers to state the general principle and leave it to each graduating school or each higher admitting school to apply the principle. The aim should be to use that Gp whether final or a projected prior Gp or to use that combination of Gp's which will most fairly and justly represent the pupil's ability.

CHAPTER XXX

SOME QUESTIONS AND ANSWERS

In this chapter will be listed some questions teachers have asked concerning the system, together with the answers.

1. *Can this marking system be simplified?* As a matter of fact the proposed system is really very simple. The person who reads it over for the first time is apt to be confused by the new terminology, and therefore may think of it as cumbersome. Teachers who have actually used this system report that it is no more complex than any other plan, and has many distinct advantages in its favor. Once a teacher has learned to operate it, the work is quite simple.

One alternative plan may be preferred by some. Defer the administration of the intelligence test until promotion time. This avoids the necessity for projection of the Gi. During the term record test scores in terms of "Number Right." Add each pupil's scores. Assign G marks on the basis of these total scores just as in the case of a single examination. This modification does not provide for the giving of marks during the term. If such marks must be given on report cards, some other plan must be used.

Another possible modification is the elimination of the plan of classification and promotion. This is possible in those systems whose policy is to promote practically all pupils, regardless of achievement, and to give no extra or double promotions.

2. *Why must the Gi's be projected?* If the Gi's as of date of test were used, the marks assigned would be too low for every month thereafter, since the Gi increases from month to month. It is not practical to project the Gi's every month. The plan of projecting them as of promotion time makes all the marks for the term strictly comparable, and simplifies the plan generally.

3. *Is it necessary in a semi-annual system to give an intelligence test at the beginning of the second semester?* No, just project the Gi another 0.5. Is it necessary to give an intelligence test every year? Yes, it is preferable to do so. The Gi, unlike the intelli-

gence quotient, changes from year to year. As a crude substitute for a new test it is possible to keep on projecting the Gi, but the further the projection the greater the error since all pupils do not grow the same amount annually.

If, for some reason, the Gi must be projected more than one year, it is recommended that the Promotion Quotient, previously described, be computed for each pupil. If a pupil's Promotion Quotient is, say, 100, add to the original Gi 1.0 for each year. If it is 110, add 1.1. If it is 140, add 1.4. If it is 80, add 0.8, and similarly for other Promotion Quotients.

4. *If the class is bright, with high Gi's, will the relatively high marks tend to give pupils a wrong impression of their achievement?* No, for such pupils generally have high achievement. We are scarcely justified in deceiving them in order to keep them satisfied with their grade classifications. There are better ways of achieving this purpose if the purpose is desirable.

5. *Will the distribution of Gi scores represent the actual distribution of achievement in every school subject?* No. The only way to get the actual distribution in any subject is to give a standard test in that subject. To administer standard tests in every subject would be expensive and marking would be confusing. The distribution of Gi's approximates the distribution in the various school subjects. Even where it doesn't there is considerable justification for the use of the marking system.

6. *Do G score marks represent the actual level of achievement indicated?* They only approximate it. A standard test in arithmetic given at the beginning of the school year may reveal that the class as a whole is somewhat below or above the norm in arithmetic, even though the intelligence may be at the norm. The fact that the distribution of marks, based as they are on the intelligence test, is typical for the grade must not blind the teacher to the possibility that the class may be above or below norm in the various subjects. They approximate actual achievement, because intelligence is a far more potent determiner of achievement than is the type of school or kind of teacher. Standard tests must supplement marks in determining actual achievement.

7. *Does this system reveal absolute differences between achievement in different subjects?* No. As shown above the G score marks do not represent the exact level of achievement, hence

differences between scores in different subjects represent relative differences and not absolute differences. If the class as a whole is above the standard test norm in arithmetic and below the norm in reading, the fact that a pupil has a mark of 5.6 in arithmetic and 6.1 in reading does not certainly mean that his achievement in arithmetic is less than his achievement in reading. However, in a class of ordinary size or larger, it is fairly safe to conclude that this pupil is better in reading than in arithmetic.

8. *Does this system reveal the inefficient teacher?* The marks will be the same for any given class regardless of who teaches it. As pointed out above, this is not a serious limitation because the actual achievement follows the intelligence pretty closely regardless of who does the teaching. This doesn't mean of course that pupils would progress just as well without a teacher! But it does mean that the differences among teachers in efficiency is not nearly so significant for growth in subject matter as the differences among pupils in intelligence. Standard tests and supplementary criteria must be used if teaching efficiency is to be determined.

9. *Does this system eliminate teacher bias?* Yes, to a large degree. Marks are assigned on an impartial basis, according to the distribution of scores. Teacher judgment does enter in marking papers to determine number right, if essay type examinations are used. However, the teacher is relieved from the necessity of determining what mark shall be assigned any one paper. The pupil's question becomes "How well did I do?" not "Did I pass?"

10. *Does the system amend an imperfect examination?* No. This plan does not improve examinations. But it does reveal how accurate examinations are, especially if two or more examinations are given which cover the same area of subject matter. In this case, if a pupil's G scores on two comparable examinations differ greatly, it means the examinations are unreliable.

11. *What is the basis for the data in Table 32 which indicates that the difference between a G_i and a subject G score, based on a forty-minute test, must be 1.0 to be significant?* In order to answer this question we must utilize statistical methods. It is not the authors' purpose to discuss the statistical concepts involved. Any book on statistical methods will serve to acquaint the

reader with the meaning of *reliability* and *Probable Error*. It is sufficient to point out here that teachers' examinations are notoriously unreliable. As a rule, the longer the examination time, or the more examinations given, the greater is the reliability of an individual's score. Assuming that teachers' examinations have a reliability coefficient of .65 and that the intelligence test has a reliability of .90, we may calculate the *Probable Error of the Difference* between Gi and G subject for various time limits (see Table 32). To be practically certain (997 times out of 1000) that the difference is a true difference, we multiply the *Probable Error of the Difference* by 4.4. Table 32 is read as follows: The Probable Error of an informal examination of 20 minutes length is .45 G, or four and one-half months. The Probable Error of the Difference between a grade score on a twenty-minute examination and a Gi, based on an intelligence test of .90 reliability is .33 G or three and one-third months. In other words, if a pupil has a Gi of 4.6 and a Ga of 4.9, the difference being .3 G or three months, the chances are equal that the difference is a true difference. The figure in the last column (1.6) shows that the difference between the Gi and the grade score on an informal examination twenty minutes long must be 1.6 to be practically certain (997 times out of 1000) that the difference is a true difference.

12. *Is there no way to avoid a few erratically high or low subject G scores due to a few abnormally high or low Gi's?* Yes. It can be done by averaging the top 10 per cent of the Gi's and assigning this average Gi to the top 10 per cent of the subject scores, and so on for the next 10 per cent, and the next, to the last 10 per cent. The scale can be coarsened still further by averaging the Gi's for the top 20 per cent, and so on. By such means it is possible to avoid using an abnormally high or low Gi due, perhaps, to an error of measurement. But such an escape may penalize or give advantage to some pupil whose Gi actually is abnormally high or low, respectively. The coarsening of the scale is suggested in the answer to the next question.

13. *Could the general technique of the grade score marking system be used in a high school or, especially, a college where it was deemed inadvisable to use the G score or age score?* Yes. What follows is a plan developed for presentation to the faculty of Teachers College, Columbia University, in an effort to prevent the

faculty from inaugurating a system of granting all degrees on the basis of comprehensive graduation examinations made up by allied departments:

THESES TO BE CONSIDERED BY THE FACULTY

Teachers College both educates and certifies. This report assumes, without prejudice, that we shall continue to do both.

A. EVALUATION OF STUDENT ACHIEVEMENT

1. Since the evaluation of achievement and the marking system, i.e., the particular method of recording achievement, are often confused, the two should be considered both separately and together.

2. In evaluating achievement, we should never lose sight of our real criterion, namely, the ability of our graduates whether teachers, principals, supervisors, psychologists, superintendents, or others to make desirable changes in students.

3. Until we can satisfactorily measure an adequate sampling of these desirable changes and also the capacity of the learner to make these changes, all our discussion of examinations, courses, balance between knowledge of the subject and technique of teaching, etc., will necessarily be academic and our conclusions will lack reasonable validity.

4. Through the collaboration of Tyler, Wrightstone, Coy, and many others criterion tests are being provided for measuring changes in students in the senior high school and possibly early college.

5. McCall and Herring in collaboration with New York City officials and representatives from the metropolitan colleges and universities are preparing similar criterion tests for elementary and junior high schools.

6. Until these programs of measurement or better ones are ready for use, the cause of education will be served best by encouraging reasonable diversity of policy and practice in the training of educators, for then we can institute causal investigations, without delay, to determine which method of training teachers is best.

7. The use of uniform achievement tests for groups of instructors will have the following objectionable effects:

(a) It will give the impression that we already know and can predict what kind of a graduate will make the most desirable changes in students, whereas the weight of the evidence to date is that we do not know and that what most think they know is wrong.

(b) It will tend to compel uniformity instead of diversity of objectives among our instructors.

(c) It will almost surely result in a greater emphasis on subject-matter learnings, whereas the whole trend of our times (rightly or wrongly) is away from subject-matter as an end and toward, what for short I shall call guidance.

(d) It will tend to take control of each course away from the instructor and lodge it in an examination committee which must surely feel incapable of discharging its obligation.

(e) It will tend to make our examinations tests of physical endurance due to the strain on the students of concentrating a decision of serious import to them within a few hours.

(f) It will be impossible to fix a time for such an examination that will not seriously penalize some students.

(g) It will tend to restrict all evaluation to paper-and-pencil tests.

8. We should, therefore, for the time being, leave the measurement of achievement to the instructor or instructors responsible for each course, guided by their fundamental conceptions of education, their experienced observation of their graduates at work in the field, and such specialists in measurement as can be made available for this purpose.

9. One commendable purpose of uniform examinations is to secure better comparability among marks, but this purpose can be realized better by altering the marking system itself.

B. MARKING SYSTEM

1. When some of our instructors give a mark of A to 60 per cent of their students and others give a mark of A to only 6 per cent, and when we consider that Professor Spence has proved that the average mental ability of all classes in Teachers College is approximately the same, we cannot but conclude that our ante-diluvian marking system is in grave need of overhauling. Marks play such a serious part in the lives of students and can be made to serve the students and the College in so many important ways, that the present laissez-faire policy toward them is inexcusable.

2. We should have and it is easily possible for us to have a marking system which:

(a) Makes marks reasonably comparable from instructor to instructor.

(b) Does not compel instructors to have similar educational objectives or evaluate achievement in a uniform manner.

(c) Does not make the unreasonable demand that instructors acting individually be charged with maintaining or raising standards of graduation.

(d) Makes it possible for us to prevent, if we wish, mere point collecting for graduation.

(e) Graduates students on the basis of an average mark fixed by the faculty acting as a whole.

(f) Permits us to raise or lower standards of graduation as we think best.

(g) Prevents the steady upward creep of marks which is the familiar characteristic of the present system both in Teachers College and all colleges.

(h) Stabilizes the meaning of a mark from class to class and year to year.

(i) Tends to prevent a student from being penalized seriously by the error or prejudice of a particular instructor.

(j) Tends to make the gap between F and D no wider socially than from B to A.

(k) Spares an instructor the embarrassment of ever failing a student.

(l) Permits a student to emphasize those courses of most value to him without endangering his degree.

(m) Provides an easy means of discouraging students from trying to carry an excessive number of points.

(n) Permits us to inform a student of his average mark at the end of each semester, if we deem this wise.

(o) Provides really usable information for degree purposes, scholarship and fellowship and loan committees, student guidance, and placement.

(p) Yields numerical marks which can be combined, averaged, correlated and the like so as to permit comparison, decision, and research.

(q) Is sufficiently flexible that it may be used to evaluate any characteristic whether physique, philosophy, personality, or knowledge of psychology.

3. A marking system that satisfies the foregoing criteria follows.

(a) Administer a test of general mental ability to all students in the College. A test of an hour or less will do. A test of musical intelligence may be used with music students, for example, in case my friend Professor Dykema feels that his students lack any general intelligence! The marks will come out about the same in the end.

(b) Make a distribution of scores for the entire College and determine percentiles for each score. This will be done in the Registrar's office.

(c) Teach anything in any way and evaluate the achievement in any manner.

(d) Report to the Registrar which students in the class fall in the top 10 per cent for achievement in that class, the second 10 per cent, and so on to the last ten.

(e) The Registrar will assign the students who are in the top 10 per cent the average of the intelligence percentile scores made by the top 10 per cent of students in that class, and similarly for the other students. These marks will be approximately comparable throughout the College. Note that it is possible for a student's intelligence percentile to be very low and his achievement percentile to be very high. Generally, however, we would not expect such reversals. The existence of an intelligence and achievement percentile for a student will permit useful guidance by some person properly qualified by his or her discretion to give it.

(f) There are many possible minor variations but the important thing is to decide for or against this basic plan.

4. The student's marks may be supplemented by records of his experiences, exhibits of products he has produced, and the like.

14. *What references, beyond those already mentioned, deal with the topics treated in Books Three and Six?* Send for an excellent annotated bibliography prepared by Segel, a small leaflet entitled *Good References on Elementary Education, Classification, Grading, Promotion*. This is Bibliography No. 39, issued by the Office of Education, United States Department of the Interior, in 1936.

-

BOOK SEVEN



PRESENTATION OF TEST RESULTS

CHAPTER XXXI

GRAPHIC METHODS

Importance of Presentation.—Other chapters may exceed this one in length, but none exceeds it in the importance of the topic considered. Recently posters appeared on New York City bill boards announcing a new play: "It Pays to Advertise." The poster showed a cackling hen leaving an egg-filled nest. For the sake of the public it is necessary to have a dignified title for this chapter. But it will not be amiss to imbed here in the privacy of the text the statement that the real title of this chapter is: It Pays to Advertise. Preceding chapters have attempted to show how the truth about conditions in the school may be discovered. Presumably these facts have not been collected to fill up files, but rather to publish in the schoolroom, at teachers' meetings, in public addresses, in school reports, or in periodicals. Presumably these facts have been collected to influence action—the action of pupils, teachers, supervisors, principals, superintendents, boards of education, or the public. Truth does not prevail through facts but through the effective presentation of facts.

There are three types of presentation in common use: the tabular, the graphic, and the linguistic. Generally speaking, that type of presentation is most significant which in the particular situation best fits the data, the purpose, the occasion, the medium of presentation, whether in an address, a published article, etc., and which best fits the kind of audience.

The graphic method is, however, generally conceded to be the best method for most situations. The graphic method is particularly effective because when graphs are properly made they are more easily and more quickly interpreted. For both these reasons, and perhaps others in addition, graphs have an intrinsic psychological appeal denied to numbers and words. It is only the unusual person whose tabular or literary skill is sufficient to overcome this inherent superiority of the graphic method. Finally, the properly constructed graph shows not only the

graph but presents tabular data and utilizes linguistic description at the same time. The graph combines most of the advantages of all three methods, and is hence a powerful instrument in the hands of intelligent educators.

Standard Graphic Methods.—The standardizations of graphic methods is just as important as the standardization of statistical procedure. In order to further a notable movement toward standardization which has already begun and in order to give the reader an introduction to graphic methods the full preliminary report of the Joint Committee on Standards for Graphic Presentation is given below.

JOINT COMMITTEE ON STANDARDS FOR GRAPHIC PRESENTATION

*Preliminary Report Published for the Purpose of Inviting Suggestions
for the Benefit of the Committee*

As a result of invitations extended by The American Society of Mechanical Engineers, a number of associations of national scope have appointed representatives on a Joint Committee on Standards for Graphic Presentation. Below are the names of the members of the committee and of the associations which have coöperated in its formation.

WILLARD C. BRINTON, *Chairman*, American Society of Mechanical Engineers.
7 East 42d Street, New York City.

LEONARD P. AYRES, *Secretary*, American Statistical Association.
130 East 22d Street, New York City.

N. A. CARLE, American Institute of Electrical Engineers.

ROBERT E. CHADDOCK, American Association for the Advancement of Science.

FREDERICK A. CLEVELAND, American Academy of Political and Social Science.

H. E. CRAMPTON, American Genetic Association.

WALTER S. GIFFORD, American Economic Association.

J. ARTHUR HARRIS, American Society of Naturalists.

H. E. HAWKES, American Mathematical Society.

JOSEPH A. HILL, United States Census Bureau.

HENRY D. HUBBARD, United States Bureau of Standards.

ROBERT H. MONTGOMERY, American Association of Public Accountants.

HENRY H. NORRIS, Society for the Promotion of Engineering Education.

ALEXANDER SMITH, American Chemical Society.

JUDD STEWART, American Institute of Mining Engineers.

WENDELL M. STRONG, Actuarial Society of America.

EDWARD L. THORNDIKE, American Psychological Association.

The committee is making a study of the methods used in different fields of endeavor for presenting statistical and quantitative data in graphic form. As civilization advances there is being brought to the attention of the average individual a constantly increasing volume of comparative figures and general data of a scientific, technical, and statistical nature. The graphic method permits the presentation of

such figures and data with a great saving of time and also with more clearness than would otherwise be obtained. If simple and convenient standards can be found and made generally known, there will be possible a more universal use of graphic methods with a consequent gain to mankind because of the greater speed and accuracy with which complex information may be imparted and interpreted.

The following are suggestions which the committee has thus far considered as representing the more generally applicable principles of elementary graphic presentation.

1. *The general arrangement of a diagram should proceed from left to right.*

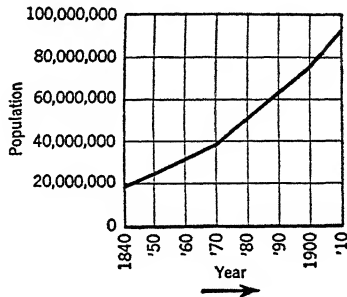


FIG. 2

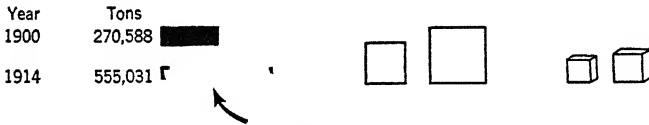


FIG. 3

2. *Where possible represent quantities by linear magnitudes as areas or volumes are more likely to be misinterpreted.*

3. *For a curve the vertical scale, whenever practicable, should be so selected that the zero line will appear on the diagram.*

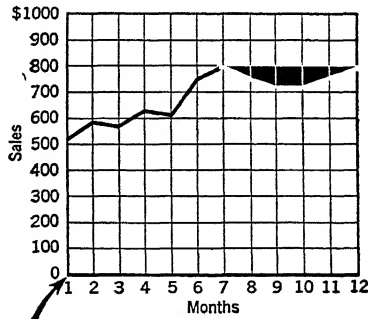


FIG. 4

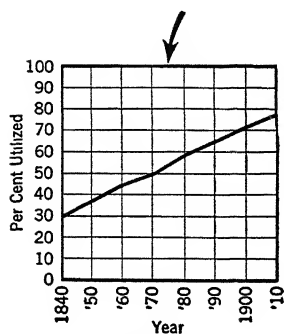


FIG. 9

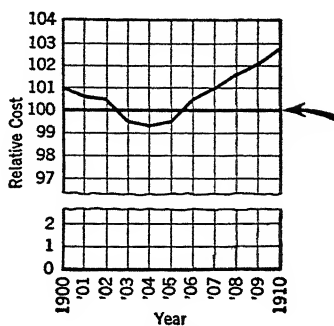


FIG. 10

6. For curves having a scale representing percentages, it is usually desirable to emphasize in some distinctive way the 100 per cent line or other line used as a basis of comparison.

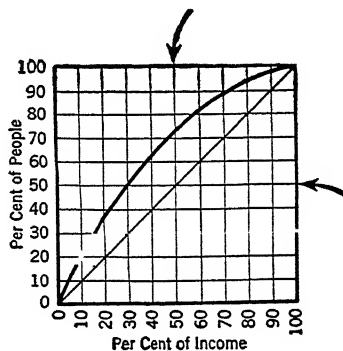


FIG. 11

7. When the scale of a diagram refers to dates, and the period represented is not a complete unit, it is better not to emphasize the first and last ordinates, since such a diagram does not represent the beginning or end of time.

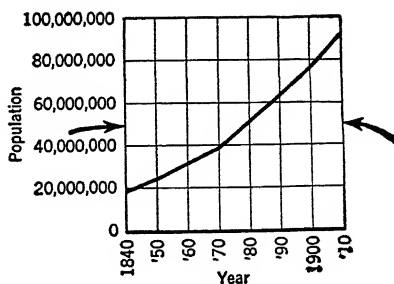


FIG. 12

8. When curves are drawn on logarithmic coordinates, the limiting lines of the diagram should each be at some power of ten on the logarithmic scales.

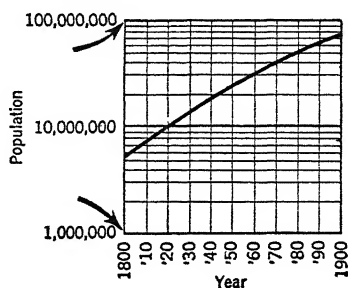


FIG. 13

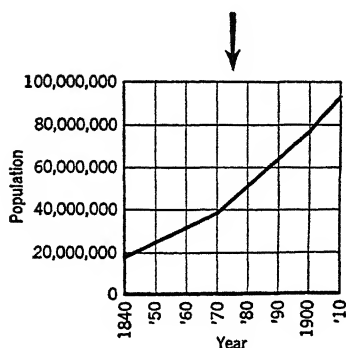


FIG. 14

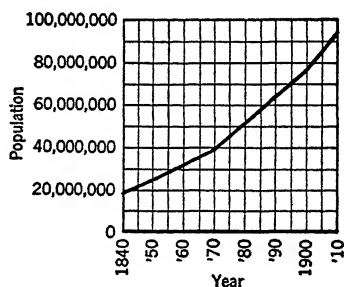


FIG. 15

9. It is advisable not to show any more coordinate lines than necessary to guide the eye in reading the diagram.

10. The curve lines of a diagram should be sharply distinguished from the ruling.

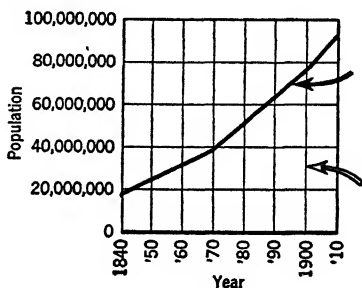


FIG. 16

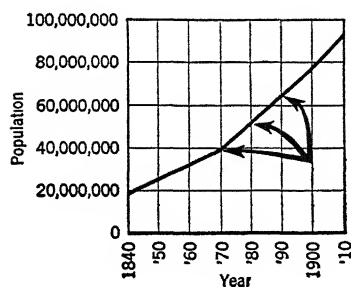


FIG. 17

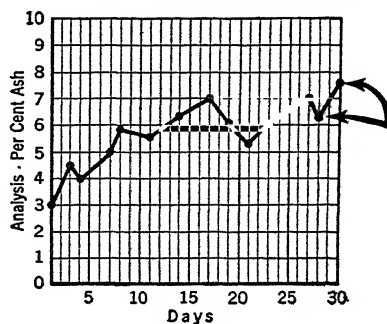


FIG. 18

11. In curves representing a series of observations, it is advisable, whenever possible, to indicate clearly on the diagram all the points representing the separate observations.

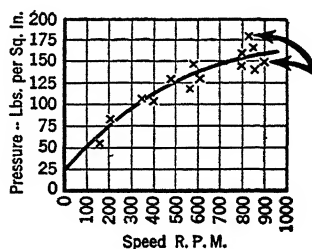


FIG. 19

12. The horizontal scale for curves should usually read from left to right and the vertical scale from bottom to top.

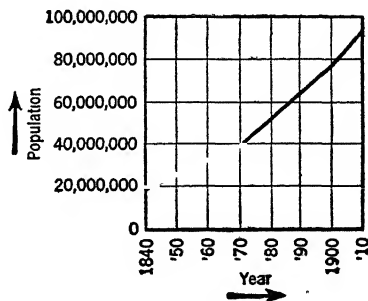


FIG. 20

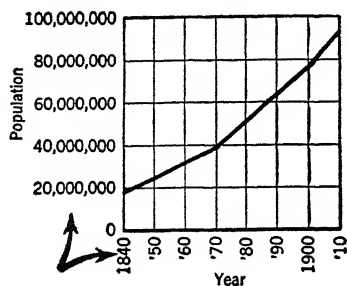


FIG. 21

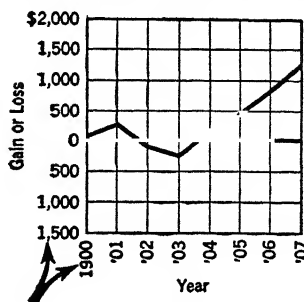


FIG. 22

13. Figures for the scales of a diagram should be placed at the left and at the bottom or along the respective axes.

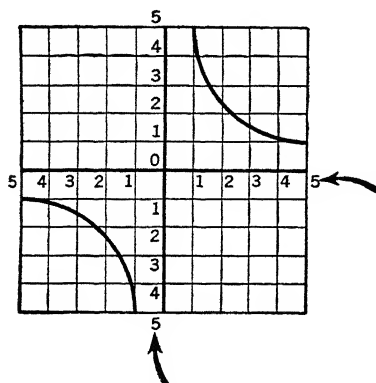


FIG. 23

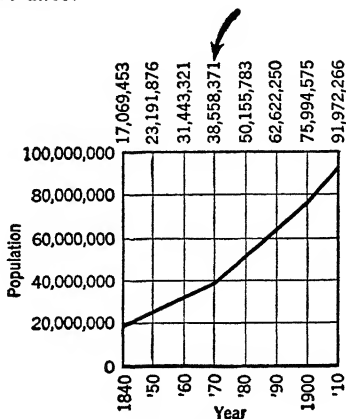


FIG. 24

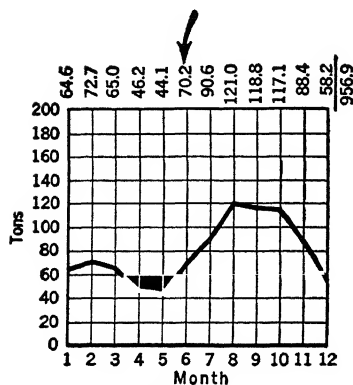


FIG. 25

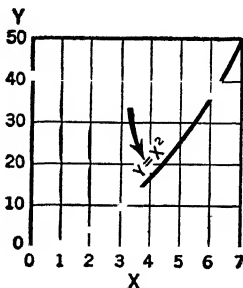


FIG. 26

14. It is often desirable to include in the diagram the numerical data or formulae represented.

15. If numerical data are not included in the diagram it is desirable to give the data in tabular form accompanying the diagram.

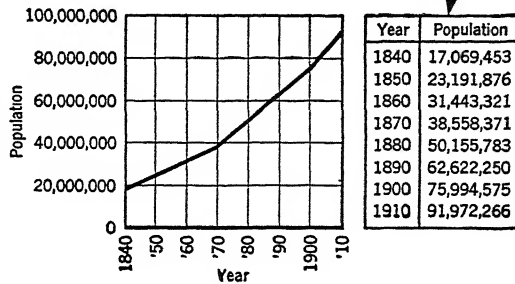


FIG. 27

16. All lettering and all figures on a diagram should be placed so as to be easily read from the base as the bottom, or from the right-hand edge of the diagram as the bottom.

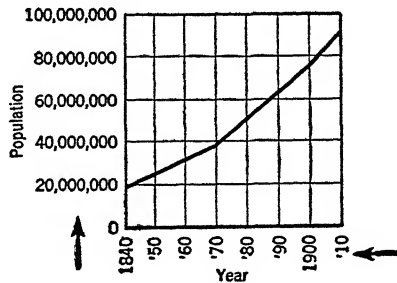
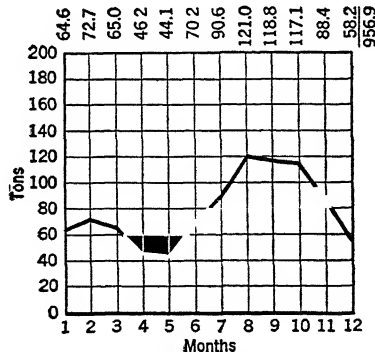


FIG. 28

17. The title of a diagram should be made as clear and complete as possible. Sub-titles or descriptions should be added if necessary to insure clearness.



Aluminum Castings Output of Plant No. 2, by Months, 1914. Output is given in short tons. Sales of Scrap Aluminum are not included.

FIG. 29

Further Principles of Graphing.—The suggestions given below do not appear in the report of the Committee on Graphic Presentation, but through the influence of Brinton's book,¹ in particular, they have become rather generally accepted as good practice. The reader is referred to his book for a further amplification and illustration of these principles.

18. *When several items are being compared the item of chief interest may be made more striking than the others.*

The most important item can be made more striking by the use of (a) capitals or red letters for the title. Thus in Fig. 3, for example, the "1914" and the "555,031" could have been printed in red, provided the year 1914 had some peculiar importance. If a principal were comparing his school with other schools he would make the title of the bar representing his own school red, or capitalize the title of his school. If, on the other hand, several schools are being compared with standard, the standard would be made red because the standard would be the most prominent item.

The important item could be made more striking by the use of (b) a solid bar for the important item and an outlined bar for the secondary items, or by the use of (c) a heavier bar or curve for the important item, or by the use of (d) a colored bar or curve for the important item. If desirable and undesirable items are being compared and more than one color is used, it has become a practice to represent the undesirable items by red and the desirable items by green.

19. *Popular features or "eye catchers" may be used to attract attention to the diagram but may not, as a rule, be an integral part of the diagram.*

If the diagram concerns the cost of producing a given unit of growth in pupils large \$'s will help to attract attention, but they should accompany the diagram and not be a part of it. That is, no attempt should be made to show the cost by the number of \$'s.

20. *Do not place captions or numbers so as to alter the length of bars or to interfere with a visual comparison of their length.*

This means that all numbers should appear at the left of the bars, unless the bars are drawn vertically, in which case the numbers may appear at the top of the bars written horizontally.

¹ Brinton, Willard C., "Graphic Methods for Presenting Facts," *The Engineering Magazine Co.*, New York, 1917, 371 pp.

Were the numbers shown to the right of the bars in Fig. 3 instead of at their left and were the tons for the lower bar a million or more the 1914 bar would be made to appear longer than it really is, due to the longer length of the numbers representing tons. The caption for each bar could also be so placed as to produce a like illusion.

21. *When a scale (time scale especially) is not consecutive indicate the gap by a wider-than-usual space interval.*

Suppose there were a column of five bars like those of Fig. 3, the top one showing the score made on a test by Grade III and the bottom one showing the score made by Grade VIII. Suppose further that there is no score or bar for Grade VII. The omission of Grade VII should be indicated by a relatively wide gap between the sixth and eighth grade bars. Otherwise the reader is likely to be misled into thinking there is a point in the elementary school where there is an exceptionally rapid growth.

22. *In graphing two or more bars or curves for comparison make their zero lines coincide.*

Anyone who has ever drawn straws to determine who shall get the only apple, or pay for the drinks, knows that he must be suspicious of the apparent length of the straws. We are never sure of our comparison until we discover the zero point of each straw. It is necessary to be equally suspicious of graphs whose zero points are not clearly revealed.

23. *Do not use a percentage curve when it is wished to show the actual amounts of increase or decrease and do not use an amount curve when it is wished to show the per cents of increase or decrease.*

Either a curve must be drawn on a logarithmic scale in order to show both amounts and per cents of change or else two graphs are required, one to show amount and one to show percentage.

As to comparable scaling, it is well to remember that of two curves plotted to the same scale and whose variability is identical, the upper curve will appear to have larger fluctuations. Statisticians are familiar with the notion that the variability of two sets of data cannot well be compared until the variability of each has been divided by the average of the data from which each variability was computed. This means that the larger the data is numerically the larger will be the amount of fluctuation, even when the percentage of variation remains constant. When it is wished to compare the fluctuation of two curves on the

same graph, one of which represents numerically small amounts and the other numerically large amounts, convert the amount curves into percentage curves and interpret in the light of the original absolute amount of each.

24. *Use a diagram which is appropriate to the data to be presented.*

What diagrams to use in a given situation is discussed below.

Types of Diagrams.—There are a bewildering variety of diagrams, some good, some bad. And there is an unlimited number of graphs which may be classed as cartoons. Such, for example, is a drawing showing which of a pupil's neural pathways are in action when he is adding, or a drawing which pictures the number of germs in the water where pupils swim or any other of the numerous pictographs. The value of such cartoons usually disappears with use and hence they are not appropriate material to consider here. A ride on a street car, or a brief study of bill boards will give enough suggestions of cartoons to use. To standardize them would be to destroy their value.

Most of the standard diagrams are variations upon a few simple types. The few types listed below will be found adequate for most persons and most purposes. If any reader plans to do a great deal of graphing he should consult some special treatise on the subject, such as Brinton's.

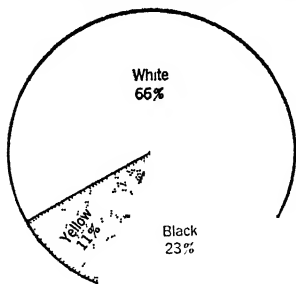


FIG. 30. Distribution by Race of the Pupils in Grades III through VIII of a Public School in an Eastern City.

Type I. The sector diagram.—Thus far in this book no illustration of the sector diagram has been printed. One is given in Fig. 30.

The construction of a sector diagram is exceedingly simple. There are 360 degrees in the circle. Sixty-six per cent of 360 degrees is 237.6 degrees. The 237.6 degrees may be roughly estimated with the eye or more accurately measured with a protractor. The other sectors are determined in a similar fashion. The diagram would be much more striking if *each* sector were colored to fit the race which the sector represents.

Type II. The bar diagram.—See, for illustration, Fig. 3.

Type III. The sectioned-bar diagram—(a) without subdivisions, and (b) with subdivisions of the component parts. The top bar of Fig. 31 illustrates the sectioned-bar diagram without subdivisions, while the entire figure illustrates the diagram with subdivisions.

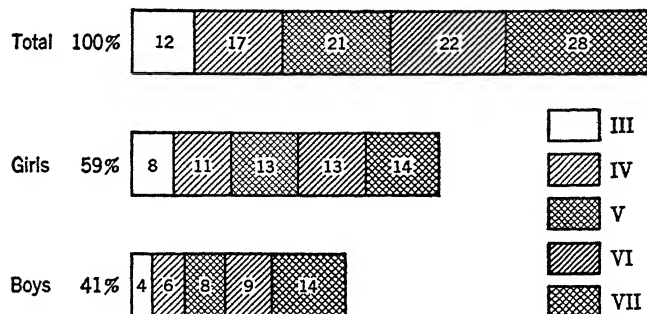


FIG. 31. The Per Cent Which the Number of Pupils in Each Grade Was of the Total Number of Pupils in All Grades Who Attained Woody's Norms According to a Random Sampling of 300 Boys and 300 Girls in a New York School.

This diagram uses such a design in each section as to make it appear distinct from the adjoining sections. This plus the sectioning makes it clear at a glance that in this particular school the per cent of pupils attaining standard gradually increases with progress through the grades. A larger percentage of girls attain standard than boys. With progress through the grades the percentage of boys who attain standard gradually increases relative to the percentage of girls. In Grade VII the boys' percentage has reached the percentage of the girls.

The unique combination of the bar and sectioned-bar diagrams shown in Fig. 32 is not only unusual but also unusually effective.

Type IV. The frequency surface.—Figure 33 may be examined as an illustration of frequency surfaces.

Type V. The curve diagram.—Numerous illustrations appear in the Report of the Joint Committee on Standards for Presenting Facts. Figure 2 may be inspected as a sample.

Practically every diagram listed above, except the sector diagram, is a bar diagram or some variation on this basic type. The sectioned-bar diagram is merely a bar diagram divided into

component parts. The frequency surface is merely a series of bar diagrams placed close together and in a vertical position. A curve diagram is merely a series of non-adjoining narrow vertical bars which are connected at the tops with a continuous line or curve.

Libraries	Sanitation	Education	Gen. Govt.	Recreation	Police Dep't	Fire Dep't	Highways	Charities
1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9
10	10	10	10	10	10	10	10	10
11	11	11	11	11	11	11	11	11

FIG. 32. Rank of Cleveland among Eighteen Cities in Expenditure for Operation and Maintenance of Schools. (After L. P. Ayres, *The Cleveland School Survey*, 1916.)

A special form of the curve diagram frequently used by mental measurers is the psychograph or mental profile. If the zero line in Fig. 8 represented the standard scores on several mental tests and the dates shown at the bottom were each the name of a mental test, then the second curve would show a sort of mental profile of an individual or group.

Selection of Diagram to Show Component Parts.—Frequently in educational measurement it is necessary to show what part each of various components is of the whole. In order to assist an audience to properly interpret certain test results it may be necessary to show what per cent of the total number of pupils in a school system belongs to the White race, Black race, etc. It may be necessary to show what per cent of pupils in Grade IV of a certain school are eight, nine, ten, or eleven years of age. It may be desirable to show how many or what per cent of the pupils, or schools, or cities make various scores on the test. All these are situations involving component parts of a whole, and require a diagram appropriate to component parts.

Perhaps the simplest of all diagrams for showing component parts is a sector diagram such as is shown in Fig. 30. The sector diagram would serve for any situation listed in the preceding paragraph.

The sectioned-bar diagram shown in Fig. 31 is an even better graph for presenting component parts. It is in almost every respect superior to the sector diagram. Visual comparisons of the components are easier. The direction of all lettering is uniform. The numerical data can be so placed that numbers and decimal points are directly under each other, so that the addition of any or all components is greatly simplified. The sector diagram is not nearly so flexible. It will satisfactorily show only one series of components. The sectioned-bar diagram will show one or more subdivisions of components. Hence, except in the situation noted below, the sectioned-bar diagram should usually be preferred to other diagrams for showing component parts.

When it is wished to show the number or per cent of pupils making various scores or who are of various ages, or in any situation where the unit is a consecutive numerical fact such as scores, ages, dates, and the like, the frequency surface is the most convenient graph, although any of the others could be used.

There are several useful variations of the frequency surface. Figure 33, for example, reveals not only the number of schools making various scores on a test but also the identity of each school making a given score. Again, a frequency surface will show subdivisions of component parts, in which case the graph really becomes a series of vertically arranged sectioned-bars.

circle of equal size showing similar facts for another school system it would be possible for the eye to roughly compare the sectors. Other graphs, however, permit an easier and more accurate comparison.

The sectioned-bar diagram will show comparisons between two series of data better than the sector diagram. If we were to place one or more graphs, showing similar data, for another school directly under the top bar of Fig. 31 the eye could, with some difficulty, compare the length of one section with the corresponding section. Comparison is made difficult by the fact that the beginning points of all corresponding sections are not directly over each other.

The frequency surface is even more useful than the sector or sectioned-bar diagrams for comparing series of components. Figure 1 illustrates such a use. Here the frequency surfaces are placed one above the other. When not more than two series of components are being compared the two frequency surfaces may be placed on the identical base line. When there are more than two surfaces on the identical base line the overlapping becomes too confusing to be useful.

The curve diagram is the most useful of all graphs. Its prevalence in the Report of the Joint Committee on Standards for Presenting Facts is a sort of index of its utility. Before finally choosing the type of diagram for presenting his data the reader will do well to go through the charts of the Joint Committee to see if some curve which he finds there may not satisfy the condition of his data. The curve is familiar to most persons; it is easily and quickly read; it is so flexible that almost any data can be presented by means of it.

The curve is particularly effective for comparing two series of similar data. Suppose we have a curve showing the progress of the medians from grade to grade of a certain school on a certain test. One or more other curves representing the grade progress of other schools on the same test may be drawn on the same diagram, thus permitting easy comparison.

Curve diagrams may also be used to compare series of components. The curve diagram could take the place of the overlapping frequency surfaces in Fig. 1. When a frequency surface is made with a series of rectangles as in Fig. 1 it is called a *histogram*. When a frequency surface is made with a continuous

curve it is called a *frequency polygon*. All that is needed to convert the histogram into a frequency polygon is to draw a continuous line which passes through the middle point of the top of each rectangle and then erase the lines which block out the rectangles. In practice, the frequency polygons are drawn directly from the data.

The curve is equally preëminent for showing the relationship between two series of data. A curve of grade progress shows the relationship which obtains between grade and score on a test. A curve of age progress shows the relationship between the age of a pupil and score on a test. Figure 19 is an illustration of how the curve type of chart may be used to give a graphic picture of correlation.

Preparation of Diagrams.—The following materials are either essential or useful in charting: appropriately ruled paper or plain paper to be ruled by the person making the chart, drawing board, T-square, decimal scale ruler, French curves, reducing glass, colored crayons, waterproof India ink of various colors, gummed letters and figures. Still other appliances would be useful but few persons outside of professional draftsmen have half the material already mentioned.

The material upon which the diagram is drawn will vary with circumstances. When test records are kept on file from year to year it will be found advisable to make diagrams for these files on ruled cards of uniform filing size. For lecturing purposes the chart may, by means of a brush, be drawn in white paint on black cambric cloth. When the paint has dried this cloth can be folded and packed into a small space in a handbag.

The charts should be drawn in harmony with the suggestions already presented. Besides this, the diagram should be neat with all lettering as plain as possible. Gummed letters and numbers may be used to produce a clearer and neater picture, if the person making the diagram is not skilled in making letters and numbers. If the diagram is intended for publication it is advisable to make the drawing larger than it will be when published, in order that in the process of reduction to printing size minor irregularities will disappear. If the graph is made just twice the printing size great care must be taken to see that every proportion of the original is exactly twice the size that is finally desired. Coördinate and other lines must be twice as wide, and

twice as far apart. Letters and numbers must be twice as high and wide and so on. These proportions may be determined by general judgment, by use of the reducing glass, or by actual measurement. Finally, the diagram should be drawn in India ink in order that it may give a clear photograph. Black, red, green, and blue India inks all photograph black. Black prints blacker than any other color; red is a close second, and the others in the order named.

Reproducing the Diagram.—If the diagram is intended for local use it may be reproduced on a hectograph or mimeograph at very little expense and with very little trouble. If this method of reproduction is used the diagram must be prepared with a special kind of ink in the former case, or on a special stencil in the latter case. Adequate instructions for this process come with these reproducing instruments. A school can ill afford to be without either a hectograph or mimeograph.

The blue print is another method of speedy and inexpensive reproduction and so is the photostat machine. The photostat machine will make direct photographic copies of diagrams. Blue-printing and photostat companies will be found in most large cities.

The stereopticon, reflectoscope, and motion picture may be considered reproducing machines. There are companies who will convert any diagram into a lantern slide whose use in connection with a stereopticon will throw the diagram on a screen. Many schools are finding the stereopticon an indispensable adjunct. There are portable stereopticons which may advantageously be taken on lecture tours. Reflectoscopes are made which will reflect a diagram directly from the paper drawing. This saves time and expense involved in having lantern slides prepared but it is not so satisfactory in other respects as the stereopticon. All are familiar with the motion picture machine.

If a diagram is published one of three methods may be employed, (a) a zinc or line cut, (b) half-tone or copper plate, or (c) Ben Day. The zinc cut is the cheapest, the half-tone next, and the Ben Day process is the most expensive. As stated before, diagrams are usually more effective when printed in color, but color printing is very expensive indeed. Before the diagram is sent to the publisher instructions as to the process and the final dimensions desired should be noted on the margin,

preferably in blue pencil since such markings do not photograph. If the process is Ben Day the shading desired for each portion of the graph should be selected from a catalog and indicated.

Those who wish for additional help on graphic and tabular presentation are referred to the following references:

Alexander, Carter, *School Statistics and Publicity*, Silver Burdette & Co., Newark, 1919.

American Society of Mechanical Engineers, *Code of Preferred Practice for Graphic Presentation—Time Series Curve Charts*, New York, 1937.

American Society of Mechanical Engineers, *Engineering and Scientific Charts for Lantern Slides*, New York, 1932.

Brinton, W. C., *Graphic Methods for Presenting Facts*, Works Management Library, New York, 1914.

Modley, Rudolph, *How to Use Pictorial Statistics*, Harper and Brothers, New York, 1937.

Walker, Helen, *Statistical Tables, Their Structure and Use*, Bureau of Publications, Teachers College, Columbia University, New York, 1937.

BOOK EIGHT

HOW TO SCALE TESTS AND COMPUTE
STATISTICAL MEASURES

CHAPTER XXXII

REFERENCE POINTS AND SCALE UNITS

Reference Point.—Whatever the measurement scoring must have some starting point—some reference point. Kalamazoo has a location, but the location is not very intelligible to anyone unfamiliar with Kalamazoo unless given some reference point or points. If we say Kalamazoo is so many degrees west longitude and so many degrees north latitude, the reference points are the line of longitude passing through Greenwich and the line of latitude corresponding to the equator. According to scientific measurement the reference point for measuring an individual's height is either the soles of the bare feet or the actual crown of the head. Whether the thing measured be distance, time, weight, courage, reading ability, or arithmetical skill, there must be a starting point for scoring.

The following drama will illustrate the need for a commonly understood reference point:

TRAGI-COMEDY OF ERRORS ACT FIRST

Railroad Station, Richmond, Va.

Enter TRAVELER, NATIVE OF BALTIMORE, NATIVE OF SAVANNAH, BOSTONIAN, and AUTHOR of this book.

TRAVELER: Is New York City farther than Philadelphia?

AUTHOR: Define your point of reference. (*Exit* AUTHOR.)

NATIVE OF BALTIMORE: Yes.

NATIVE OF SAVANNAH: Yes.

BOSTONIAN; No! (*Exit* BOSTONIAN.)

TRAVELER: How much farther is New York City than Philadelphia?

NATIVE OF BALTIMORE: About twice.

NATIVE OF SAVANNAH: About one-tenth!

The End

Scientists soon discovered that scientific progress was handicapped by the fact that different individuals were using different reference points when measuring temperature. Finally after long wasteful delays two competing reference points have been adopted, one which places the zero point of the temperature

scale 32 degrees below the freezing point of water and one which locates it at the freezing point of water. In similar manner scientists agreed to make the zero for the height of land forms the sea level. They could have made it the center of the earth or the base of the Acropolis. In the measurement of many things in life then there is no one point divinely called to be zero. Convenient zero points have been proposed, debated, and arbitrarily adopted.

Mental measurers have for years been searching for an appropriate reference point or points. The tendency has been to search for some absolute zero point for the trait being measured. This has resulted in a different zero point for each scale made. If the process continues we shall have hundreds of zero points each of which is extremely nebulous, and no one of which is generally accepted. The resulting confusion would enormously handicap the development of mental measurement.

We have had not only a different reference point for each test, but different methods of locating this point. *First*, the reference point on unscaled tests is just no score on the material of the particular test. *Second*, the reference point on certain scales is a zero point guessed at by the author of the scale. *Third*, the reference point on other scales, particularly judgment scales, is the median judgment of judges as to the location of zero merit in composition, handwriting, art, etc. *Fourth*, the reference point on other scales is a zero point located by the use of the per cent of pupils in some early grade who make no score on very easy material. *Fifth*, the reference point for other scales is 3 S.D. (see Chapter XXXV) below the mean of the group for whom the test was devised. *Sixth*, the reference point on another scale is simply the lowest score made. Still other methods of locating reference points have been used.

Since few mental measurers agree as to the best method of locating a zero point, since few agree as to just what the zero for reading or any other mental trait is, since any such point if actually found is bound to be relatively invisible and hence more or less valueless as an aid to the proper interpretation of scores, since prevailing methods of locating zero are certain to produce as many different points as there are scales, and since this last must inevitably result in general confusion, this book proposes that three reference points be arbitrarily adopted for all tests which are to be used in the elementary school. It is recom-

mended that these reference points be the beginning point of the kindergarten for the grade scale, the time of birth for the age scale, and 5 S.D. below the mean performance of children between the ages of 12.0 and 13.0 for the T scale.

Unit of Measurement.—Just as all measurement requires some reference point so all measurement requires some unit. The reference point for a mountain is sea level. Its height above this reference point is expressed in terms of a certain measuring unit called a *foot*. The reference point for measuring time is the birth of Christ, or January 1st, or 12 M., and the units are centuries, years, days, hours, minutes, and seconds.

The variety of reference points is almost equaled by the variety of units for mental measurement. Thorndike and his students have used some function of variability as a unit and this is admirable. They have used the variability of a grade, which is not so admirable. Many forces are at work such as reorganizations of grade systems, improvements in classification, and the like, which are bound to profoundly alter, in a relatively short time, all scale values and the significance of the scale units employed. Any unit based upon such an artificial and ephemeral group as a grade lacks the necessary permanence. They have used values based upon the variability of several grades and have combined these values through an elaborate procedure of weighting values and determining inter-grade intervals. This procedure has the merit of giving temporarily reliable results, but the whole procedure is altogether too laborious for it to be generally used. Furthermore, the values when pooled for several grades with intricate weightings cease to be interpretable. The only sort of variability which has much meaning is the variability of some one defined group. Even so this group of scientific workers has done more to further the cause of accurate scale construction than any other group in the world.

The other high point in scale construction began with Binet and Simon and culminated in the *Stanford Revision of the Binet-Simon Scale* by Terman. This line of development has been popular rather than technical. Its reference point has been the time of birth, and its unit of measurement has been one year of growth or some subdivision thereof. These are reference points and scoring units which all can understand. They utilize chronological age, one of the most abiding features of human life.

There are, however, some very serious objections to this unit of measurement. While a permanent one, it is not equal in the truest sense of the word, at all points on the scale. A fact, now taken for granted, is that the interval between 8 and 9 years of age is larger than the interval between 14 and 15, in the case of intelligence and probably for many school traits as well. Furthermore, in the case of certain mental traits, the units become of zero size beyond about sixteen years of age. In abilities where a loss occurs, after instruction in the elementary school ceases, the age unit may be actually less than zero, i.e., negative. Finally, because of the late entrance into school of some pupils and because of the disappearance into the social medium of a goodly per cent of the graduates and over-compulsory-school-age pupils of the elementary school, it becomes difficult, if not impossible, to build up a scale below an age of 8 years and above an age of 12 years. This means that such a restricted scale cannot satisfactorily score a very poor or very able pupil. To accurately extend the scale so it will measure these individuals requires that a test be previously scaled beyond these points by some other method of scaling. Lending itself as it does to easy interpretation and to the ready computation of quotients, the age unit is deservedly popular.

The unit employed by the T scale, namely one-tenth S.D. of twelve-year-old children, has long been used by careful mental measurers to compare pupils with other pupils in their own age group. This unit is equal at all points on the difficulty scale, which is the chief characteristic of the unit employed by Thorndike and his students. It is based upon chronological age which is the chief characteristic of the work of Terman and his predecessors. It is a function of the variability of a defined group and a group which is easily located. A scale which uses this unit reaches as low and as high as the ordinary requirements of practical measurement. Special extension at the top or bottom is a simple process. And not of least importance is the fact that the construction of the scale which employs this unit is not particularly laborious. In sum the proposed unit combines most of the virtues and eliminates most of the defects of the chief contemporary methods of constructing mental scales. In a certain sense it unites the two great lines of scale development. Since the two greatest contemporary exponents of these merging

methods are Thorndike for the one and Terman for the other, it is a tribute to their genius to call the proposed unit, namely one-tenth S.D. of unselected twelve-year-old children, a Thorndike-Terman, or, for brevity, a T.

The unit employed by the grade scale is the amount of growth between any two adjoining grades composed of typical pupils. Because this unit is so useful in classifying pupils and is so readily understood by all teachers, it has become the most extensively used of all test units, even though the unit lacks both permanence and equality.

In the grade scale, the mean number of points made on the test in question by a typical third grade at the beginning of the third grade is assigned a score of 3.0, and any pupil gets a score of 3.0 if he makes the number of points corresponding to it. The mean number of points made by the fourth grade at the beginning of the fourth grade is assigned a value of 4.0 and so on. Intermediate values can be determined by interpolation.

Another unit, once very popular, but now used mostly in connection with tests for adults is the percentile.

In the percentile scale, the smallest number of points made on the test in question by any pupil of the group used as the basis for scaling is scored zero, the number of points below which are one per cent of the pupils is scored 1, the number of points below which are two per cent of the pupils is called 2, and so on to the highest number of points made by any pupil which is scored 100.

This method assumes that the difference in ability between a pupil who makes a zero-percentile score and a pupil who makes a 10-percentile score is the same as the difference between a pupil who makes a 40-percentile score and a 50-percentile score. It is rather generally conceded, however, that the former difference is actually much greater than the latter difference, and that therefore the units are not equal in the truest sense in all parts of the scale.

The newest and perhaps the most promising of all these units is one proposed by Courtis and called by him an isochron—equal time unit.

He determines and draws the growth curve for some tested ability, letting the base line of the graph represent years and months. He locates the point where the growth curve ceases to

rise any more—where the physiological limit for that ability has been reached. He finds the point on the base line which is directly below this maturation point. He calls this point on the base line or time line 100. Next he locates the point on the time line where the ability is at zero and calls this point 0. The point on the time line half-way between 0 and 100 he calls 50. The quarter points are called 25 and 75 and so on for finer scoring. To find any pupil's isochron score, all that is required is to locate that score on the vertical axis of the graph, draw a horizontal line from it to the growth curve, and then a vertical line down to the horizontal axis, and read the isochron value. Or a table may be constructed by doing this once for all possible scores. Courtis claims to have established that isochron scores are truly comparable from one mental trait to another, from the mental to the physical, from human life to animal or plant life, in fact throughout all the realms of organic life, the curve of growth being the same from pumpkin seed to pumpkin as the birth of the infant to his intellectual maturity at manhood.

Two other units that have been proposed but which have not yet come into general use are Van Wagenen's C unit (see the manual accompanying *Unit Scales of Attainment* listed in Table 1) and Thorndike's *Absolute Scale Unit* (see reference at the end of Chapter XIV, 1).

CHAPTER XXXIII

SCALES AND THEIR CONSTRUCTION

Methods of Combining Units.—The need for equality of units and a good method of combining them is shown in Table 43.

TABLE 43
SHOWING THE NEED FOR EQUAL UNITS OF MEASUREMENT
(R=RIGHT. W=WRONG)

TEST ITEMS	1	2	3	4	5	6	7	8	Score
Difficulty	1	2	3	3.1	3.2	3.3	3.7	4	
Pupil A	R	R	R	W	W	W	W	W	3
Pupil B	R	R	R	R	R	R	W	W	6

Pupil A solves three problems correctly. His unscaled score is, therefore, 3, as shown in the table. Pupil B solves six problems. His unscaled score is 6, as shown. Employing unscaled units of measurement in this manner makes Pupil B appear much more competent in comparison with Pupil A than he really is. The difficulty of solving six problems, namely 3.3, is only slightly above the difficulty of solving three problems, namely 3. A very small superiority of ability on the part of Pupil B enabled him to double his unscaled score. The use of equal units of difficulty gives Pupil A a score of 3 and Pupil B a score of 3.3.

But to call a pupil's score the scale value of the most difficult test element done correctly is subject to the objection that pupils are unable frequently to do correctly test elements of less scale value. Depending as it does upon a single test element, the score would also be rather unreliable. The only satisfactory procedure thus far devised to meet these two difficulties is too complicated for practical use.

On the other hand, to call a pupil's score the sum of the scale values of the test elements done correctly is somewhat laborious, and, in addition, is subject to the criticism that a score yielded by such a cumulative total shows the number of units of work done rather than the ability level reached. It would be like

measuring a man's lifting strength by adding the weights of a variety of weights lifted. The preceding simple-total procedure appears preferable. The man's lifting strength, according to the simple-total procedure, would be the weight of the heaviest object the man could barely lift.

For the foregoing reasons, the drift is away from the scaling of the separate test elements, except in a rough way for the purpose of arranging test elements in an approximate order of difficulty. The drift is in the direction of scaling, i.e., determining the difficulty of doing correctly a given number of the test elements in a given test. Stated differently, the drift is toward scaling total scores instead of test elements.

The grade, T, percentile, age, and isochron scales, all scale total scores.

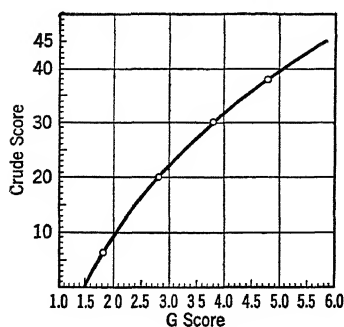
Construction of Grade Scale.—G tables are not available for some standard tests. It therefore becomes necessary to construct them, though it is better to avoid tests which are not provided with a G table. Two methods may be used: the graph method and the interpolation method.

The graph method is more accurate and will be described first. The procedure will be illustrated in the following paragraphs in

the case of the *Haggerty Reading Examination, Sigma 1*.¹

Step 1.—Take a sheet of coordinate or cross-section paper and draw a vertical line near the left-hand margin. Lay off on this line a scale of crude scores from zero to the maximum score obtainable on the test. In the *Haggerty Reading Examination, Sigma 1*, this scale extends from 0 to 45 (see Fig. 34).

FIG. 34. Graphic Method of Constructing G Table (*Haggerty Reading Examination, Sigma 1*, Used as a Sample).



Step 2.—From the base of the vertical line draw a horizontal line to the right. On this line lay

off a scale of grade scores from 0 or 1.0 as high as is needed. In our sample test, this scale extends from 1.0 to 6.0.

¹ Published by World Book Company, Yonkers. Since this account was prepared, the World Book Company has published a G Table for this examination.

Step 3.—In the manual of directions for the test, find the grade norms. For the *Haggerty Reading Examination, Sigma 1*, they are as follows:

Grade	Total Crude Score
1	6
2	20
3	30
4	38

Step 4.—In the manual of directions for the test, find the month or time of the year for which the norms are intended. The manual for the *Haggerty Reading Examination* states that the tests on which the norms are based were given in April and May. The best single date is, therefore, May 1. Let us represent the grade level on May 1 by adding the decimal 0.8 to each grade. We then have the following table of norms:

Grade	Total Crude Score
1.8	6
2.8	20
3.8	30
4.8	38

Many test manuals give norms as of October 1. In this case the Grade column would read 1.1, 2.1, 3.1, and so on. Other test manuals do not indicate the time of year for which the norms are intended. This is frequently true when norms are given by half-grades, that is, for 1L, 1H, etc. In such cases assume the norm to be intended for the middle of the grade or half-grade. For example, assume 1L to mean 1.3; 1H, 1.8, etc.

Step 5.—Plot the grade norms thus obtained on the cross-section paper. For example, erect an imaginary vertical line at 1.8, and an imaginary horizontal line at 6. Place a dot at the point where the lines intersect. Similarly, plot 2.8 and 20, 3.8 and 30, and so on.

Step 6.—Draw the straight or curved line that will best fit the plotted scores. This line should not be zigzag. It must not resemble the old rail fence. In the *Haggerty Reading Examination*, it was possible to make the curve pass through every dot. When this is not possible, try to obtain a knife-edge balance; that is, let the dots on one side be as numerous and as far from the line as the dots on the other side.

Step 7.—Construct the G Table. In the first column record all possible crude scores, beginning with zero. Read the correspond-

TABLE 44
G TABLE FOR HAGGERTY READING EXAMINATION SIGMA I

CRUDE SCORE	G SCORE	CRUDE SCORE	G SCORE
0	1.5	23	3.1
1	1.6	24	3.2
2	1.6	25	3.3
3	1.6	26	3.4
4	1.7	27	3.5
5	1.7	28	3.6
6	1.8	29	3.7
7	1.9	30	3.8
8	1.9	31	3.9
9	2.0	32	4.0
10	2.1	33	4.1
11	2.1	34	4.2
12	2.2	35	4.4
13	2.3	36	4.5
14	2.3	37	4.7
15	2.4	38	4.8
16	2.5	39	5.0
17	2.5	40	5.1
18	2.6	41	5.2
19	2.7	42	5.2
20	2.8	43	5.3
21	2.9	44	5.5
22	3.0	45	5.7

ing G scores from the graph. Thus we read on the curve opposite zero, 1.5; opposite 1, 1.55, or 1.6; opposite 2, 1.6, etc. The entire table is constructed in this way. In Table 44 it will be observed that crude scores of 39 to 45 convert to G scores of 5.0 to 5.7. These scores at the upper limit of the test are apt to be incorrect, since the test is designed for Grades I to III.

Another method for the construction of a G Table is the interpolation method. Suppose that norms as of October 1 are given. Each norm is subtracted from the norm of the grade above. The differences represent achievement during the several grades. It is assumed that achievement is in equal steps. Hence the differences are divided by 10. The result is the monthly increment. If this monthly increment is added to the October 1 norm, we have the November 1 norm, which transmutes to a G score of 3.2, in the case of a third grade. The increment is added for each succeeding month.

Construction of T Scale.—The detailed process of constructing a T scale is illustrated in Table 45. The second column

TABLE 45
SHOWING HOW TO SCALE TOTAL SCORES

TOTAL NUMBER OF QUESTIONS CORRECT	NUMBER OF TWELVE-YEAR- OLD PUPILS	NUMBER EXCEEDING PLUS HALF THOSE REACHING	PER CENT EXCEEDING PLUS HALF THOSE REACHING	SCALE SCORE
0	3	498.5	99.7	23
1	1	496.5	99.3	25
2	2	495.0	99.0	27
3	1	493.5	98.7	28
4	2	492.0	98.4	29
5	2	490.0	98.0	29
6	2	488.0	97.6	30
7	2	486.0	97.2	31
8	4	483.0	96.6	32
9	2	480.0	96.0	32
10	2	478.0	95.6	33
11	10	472.0	94.4	34
12	3	465.5	93.1	35
13	8	460.0	92.0	36
14	8	452.0	90.4	37
15	13	441.5	88.3	38
16	15	427.5	85.5	39
17	18	411.0	82.2	41
18	28	388.0	77.6	42
19	26	361.0	72.2	44
20	34	331.0	66.2	46
21	40	294.0	58.8	48
22	40	254.0	50.8	50
23	41	213.5	42.7	52
24	37	174.5	34.9	54
25	31	140.5	28.1	56
26	35	107.5	21.5	58
27	24	78.0	15.6	60
28	26	53.0	10.6	62
29	21	29.5	5.9	66
30	14	12.0	2.4	70
31	3	3.5	0.7	75
32	1	1.5	0.3	78
33	1	0.5	0.1	81
34	0			85
35	0			90

shows the number of unselected twelve-year-old children answering correctly the number of questions indicated in the first column. It is recommended that unselected twelve-year-olds (12.0-13.0) be used for scaling tests which are to be used generally. If any other age is used it should be indicated by a subscript, thus, T₁₁ or T₁₃ or T₁₆ in all publications. For experimental purposes the experimenter may use the group or groups upon which he is experimenting. The third column shows the number of pupils exceeding plus half those reaching each total number of questions correct. Thus the number of pupils exceeding 33 is 0. Half those reaching 33 is 0.5. The sum of 0 and 0.5 is 0.5 as shown in the third column. The number exceeding 32 is 1. Half those reaching 32 is 0.5. The sum of 1 and 0.5 is 1.5 as shown. The number exceeding 31 is 2. Half those reaching 31 is 1.5. The sum of 2 and 1.5 is 3.5, and similarly for other results shown in the third column. Since there are 500 pupils in the group used for scaling, the fourth column is obtained by dividing the results in the third column by 500 and by expressing the quotients as per cents. The fifth column gives the T score, and is found by converting the per cents in the fourth column by means of Table 46. Thus a per cent of 99.7 corresponds to 22.5 or, for convenience, 23.

The first column in Table 45 shows the number of test elements done correctly, where each element done counts one point. The process of scaling is the same whether each element done correctly gives a credit or penalty of one point, two points, or any number of points, or a different number of points for different elements. Thus in scoring compositions, the scorer may wish to penalize one point for each error in punctuation, and two points for each error in choice of words. If penalties instead of credits are used the first column should be inverted, i.e., large quantities should appear at the top.

Increasing the Range of a T Scale.—The width of range of a T scale based on twelve-year-olds is much wider than the inexperienced individual would suspect. In a continuous function like reading, such a T scale will measure first-grade pupils and most university students. Of course, these extreme measurements will be more unreliable than those nearer the center of the distribution for twelve-year-olds. In certain non-continuously-taught functions like algebra, or even in functions like reading, it

may be desirable to widen the range that twelve-year-olds would yield. This can be done by repeating the process shown in Table 45 for, say, nine-year-olds and sixteen-year-olds who are in high

TABLE 46

SHOWING THE S.D. DISTANCE OF A GIVEN PER CENT ABOVE ZERO. EACH S.D. VALUE IS MULTIPLIED BY 10 TO ELIMINATE DECIMALS. THE ZERO POINT IS 5 S.D. BELOW THE MEAN. S.D. VALUE EQUALS T.

S.D. VALUE	P.R. CENT	S.D. VALUE	P.R. CENT	S.D. VALUE	P.R. CENT	S.D. VALUE	P.R. CENT
0	99.999971	25	99.38	50	50.00	75	0.52
0.5	99.999963	25.5	99.29	50.5	48.01	75.5	0.51
1	99.999952	26	99.18	51	46.02	76	0.47
1.5	99.999938	26.5	99.06	51.5	44.04	76.5	0.40
2	99.99992	27	98.93	52	42.07	77	0.35
2.5	99.99990	27.5	98.78	52.5	40.13	77.5	0.30
3	99.99987	28	98.61	53	38.21	78	0.26
3.5	99.99983	28.5	98.42	53.5	36.32	78.5	0.22
4	99.99979	29	98.21	54	34.46	79	0.19
4.5	99.99973	29.5	97.98	54.5	32.64	79.5	0.16
5	99.99966	30	97.72	55	30.85	80	0.13
5.5	99.99957	30.5	97.44	55.5	29.12	80.5	0.11
6	99.99946	31	97.13	56	27.43	81	0.097
6.5	99.99932	31.5	96.78	56.5	25.78	81.5	0.082
7	99.99915	32	96.41	57	24.20	82	0.069
7.5	99.9989	32.5	95.99	57.5	22.66	82.5	0.058
8	99.9987	33	95.54	58	21.19	83	0.048
8.5	99.9983	33.5	95.05	58.5	19.77	83.5	0.040
9	99.9979	34	94.52	59	18.41	84	0.034
9.5	99.9974	34.5	93.94	59.5	17.11	84.5	0.028
10	99.9968	35	93.32	60	15.87	85	0.023
10.5	99.9961	35.5	92.65	60.5	14.69	85.5	0.019
11	99.9952	36	91.92	61	13.57	86	0.016
11.5	99.9941	36.5	91.15	61.5	12.51	86.5	0.013
12	99.9928	37	90.32	62	11.51	87	0.011
12.5	99.9912	37.5	89.44	62.5	10.56	87.5	0.009
13	99.989	38	88.49	63	9.68	88	0.007
13.5	99.987	38.5	87.49	63.5	8.85	88.5	0.0059
14	99.984	39	86.43	64	8.08	89	0.0048
14.5	99.981	39.5	85.31	64.5	7.35	89.5	0.0039
15	99.977	40	84.13	65	6.68	90	0.0032
15.5	99.972	40.5	82.89	65.5	6.06	90.5	0.0026
16	99.966	41	81.59	66	5.48	91	0.0021
16.5	99.960	41.5	80.23	66.5	4.95	91.5	0.0017
17	99.952	42	78.81	67	4.46	92	0.0013
17.5	99.942	42.5	77.34	67.5	4.01	92.5	0.0011
18	99.931	43	75.80	68	3.59	93	0.0009
18.5	99.918	43.5	74.22	68.5	3.22	93.5	0.0007
19	99.903	44	72.57	69	2.87	94	0.0005
19.5	99.886	44.5	70.88	69.5	2.56	94.5	0.00043

TABLE 46—*Continued*

S.D. VALUE	PER CENT	S.D. VALUE	PER CENT	S.D. VALUE	PER CENT	S.D. VALUE	PER CENT
20	99.865	45	69.15	70	2.28	95	0.00034
20.5	99.84	45.5	67.36	70.5	2.02	95.5	0.00027
21	99.81	46	65.54	71	1.79	96	0.00021
21.5	99.78	46.5	63.68	71.5	1.58	96.5	0.00017
22	99.74	47	61.79	72	1.39	97	0.00013
22.5	99.70	47.5	59.87	72.5	1.22	97.5	0.00010
23	99.65	48	57.93	73	1.07	98	0.00008
23.5	99.60	48.5	55.96	73.5	0.94	98.5	0.000062
24	99.53	49	53.98	74	0.82	99	0.000048
24.5	99.46	49.5	51.99	74.5	0.71	99.5	0.000037
						100	0.000029

school and elementary school, or just in high school, and by combining the results obtained with the results for twelve-year-olds. Table 47 illustrates a rough method for effecting such a combination. Simpler still, the scale may be extended graphically by extrapolation.

TABLE 47
SHOWING HOW TO WIDEN THE RANGE OF A T SCALE

PROBLEMS CORRECT	T9	T	T16	FINAL T SCALE
0	32			22
1	36			26
2	40			30
3	43	33		33
4	46	35		35
5	48	38		38
6	50	40		40
7	52	43		43
8	54	45	34	45
9	58	48	37	48
10	61	50	40	50
11	65	53	42	53
12	70	56	45	56
13		59	47	59
14		63	50	63
15		67	53	67
16		71	56	71
17		75	60	75
18		80	65	80
19			70	85
20			76	91

Construction of Percentile Scale.—If the first per cent in the fourth column of Table 45 were subtracted from 100, the remainder would be the percentile score to which 0 questions correct is entitled, and similarly for other per cents in the column. Thus a pupil with a score of 1 receives a percentile score of 0.7. A pupil with a score of 26 receives a percentile score of 78.5.

Construction of Age Scale.—In the case of the age scale, the mean number of points made on the test in question by unselected eight-year-old pupils is scored 8.5. The mean number of points made by nine-year-olds is scored 9.5, and so on. Intermediate scores are given also.

The process is, thus, very simple provided mean scores made by *unselected* pupils in the various age groups are available. But practically it is very difficult to secure such. Table 48 pictures what is generally available when a test has been given from Grades II or III up through VIII. The following procedure is recommended for determining a mean, for each age group, that is corrected roughly for selection.

1. Construct age distributions like those shown in Table 48.
2. Compute the total number of pupils for each age, and write it below the appropriate frequency column, as shown in Table 48.
3. Construct a T scale on the basis of the twelve-year-olds, and write the T-scale value in the second column, as shown in Table 48.
4. Compute half the total number of pupils for the youngest age. The half-sum or one-half the seven-year-olds in Table 48 is one-half of 35, i.e., 17.5 pupils.
5. Begin at the bottom of the frequency column for the youngest age, and add up the frequencies until the next addition or frequency will *exceed* the half-sum. Take half of this next frequency and add it to the total up to that frequency. The result will be the familiar "number exceeding plus half those reaching" the T score shown at the left. To illustrate, the half-sum for seven-year-olds is 17.5. Counting up the seven-year-old frequency column, we have $1 + 0 + 3 + 1 + 2 + 0 + 2 + 1 + 4 + 2 + (2 \div 2) = 17$. This 17 is the number exceeding plus half those reaching a T score of 34.
6. Divide the "number exceeding plus half those reaching" found in (5) by the total number of twelve-year-olds. The total number of twelve-year-olds is 500, so $17 \div 500$ gives 3.4 per cent.

7. Convert this per cent into a T score by means of Table 46. This gives 68, as shown at the bottom of Table 48. Had all seven-year-olds been tested, and had a T7 scale been constructed, the T score for 11 questions correct would have been approximately 68.

The procedure outlined above assumes that there are no seven-year-olds who read better than the better half of the 35 pupils tested. This assumption is a reasonable one, and becomes more reasonable for ages 8, 9, 10, and 11. The procedure also assumes that, since there are 500 unselected twelve-year-olds, there must be an equal number of seven-year-olds in the lower grades or community.

8. Tabulate the corresponding T score for twelve-year-olds beneath this T score for seven years. Thus, Table 48 shows 34 beneath 68.

9. Subtract the T7 score from the T12 score. The remainder is 34 and is negative, as shown in Table 48.

10. Repeat Steps 4, 5, 6, 7, 8, and 9 for all other ages up to 12. The B correction for twelve-year-olds will be zero. To give another illustration, the arithmetic of these steps for eleven-year-olds follows. (a) $426 \div 2 = 213$. (b) $1 + 0 + 6 + 4 + 3 + 13 + 16 + 16 + 22 + 29 + 32 + 40 + (35 \div 2) = 199.5$. (c) $199.5 \div 500 = 39.9$ per cent. (d) 39.9 per cent $= 52.5$ T11. (e) $48 - 52.5 = -4.5$.

11. The computation of B corrections for ages above 12 is closely similar to that for ages below 12. The only difference is that, for ages above 12, account must be taken of the fact that the better readers rather than the poorer readers are missing from Table 48. This can be done by determining the number of missing pupils, and then by adding this number in, after adding up the frequency column to find the half-sum. For thirteen-year-olds the number of pupils missing is $500 - 452$, i.e., 48. Note how this 48 is utilized in the following computations for thirteen-year-olds. (a) $452 \div 2 = 226$. (b) $2 + 1 + 5 + 11 + 19 + 25 + 24 + 39 + 46 + 42 + (42 \div 2) = 235$. (c) $235 + 48 = 283$. (d) $283 \div 500 = 56.6$ per cent. (e) 56.6 per cent $= 48.5$ T13. (f) $52 - 48.5 = +3.5$.

The B corrections for all the ages are shown in Table 48. The corrections for ages 7, 16, and 17 are quite unreliable due to the small number of cases.

TABLE 48

SHOWING THE NUMBER OF PUPILS FOR THE AGES 7 TO 17 ANSWERING CORRECTLY THE NUMBER OF QUESTIONS INDICATED IN THE FIRST COLUMN AND HENCE MAKING THE SCALE SCORES INDICATED IN THE SECOND COLUMN

No. OF QUESTIONS	SCALE SCORE	7	8	9	10	11	12	13	14	15	16	17
0	23	1	3	1	2	1	3	5				
1	25	2	3	3	4	1	1	0				
2	27	2	3	2	1	1	2	0	1			
3	28	3	0	6	3	1	1	0	0			
4	29	0	5	5	5	1	2	0	0	2		
5	29	2	5	9	6	1	2	1	2	0	1	
6	30	2	6	6	5	1	2	2	1	0	0	
7	31	0	10	6	3	5	2	2	0	0	0	
8	32	1	8	9	6	4	4	0	1	0	0	
9	32	2	10	5	5	2	2	1	0	0	0	
10	33	2	6	15	8	6	2	3	2	0	0	
11	34	2	11	20	5	4	10	1	0	1	0	
12	35	2	9	21	12	3	3	6	2	1	0	
13	36	4	14	25	12	4	8	3	1	1	0	
14	37	1	12	23	17	12	8	4	1	3	0	
15	38	2	13	21	25	15	13	12	5	2	0	
16	39	0	17	25	23	22	15	6	4	3	0	
17	41	2	17	34	24	31	18	14	4	4	0	
18	42	1	5	20	25	20	28	19	11	5	1	
19	44	3	3	20	27	32	26	26	21	3	0	
20	46	0	4	22	33	42	34	26	19	5	1	
21	48	1	4	18	25	35	40	32	28	10	2	
22	50		2	6	30	40	40	35	25	6	1	
23	52		2	6	27	32	41	42	24	9	2	
24	54		1	8	16	29	37	42	38	8	1	
25	56			3	17	22	31	46	24	16	2	
26	58			6	9	16	35	39	23	18	1	2
27	60			0	11	16	24	24	17	8	2	
28	62			2	3	13	26	25	23	5	1	
29	66				7	3	21	19	12	5	0	
30	70				2	4	14	11	7	2	1	
31	75				1	6	3	5	4	1		
32	78					0	1	1	3			
33	81					1	1	2				
34	85											
35	90											
Total Pupils		35	173	347	399	426	500	452	303	118	16	2
T7-17 Scale Score . .		68	59.5	53.5	53	52.5	50	48.5	44	38	28	21
T12 Scale Score . . .		34	36.0	38.0	44	48	50	52.0	52	54	52	58
B Correction		-34	-23.5	-15.5	-9	-4.5	0	+3.5	+8	+16	+24	+37
Mean		16	26.5	34.5	41	45.5	50	53.5	58	66	74	87
No. of Questions . .		?	1.8	11.5	17	19.8	22	23.8	26	29	30.8	34.4

12. Add each correction to 50 to get the mean T score that most likely would have been made had all pupils of a given age been tested. Thus 50 plus (-34) equals 16, the mean T score for unselected seven-year-olds.

13. Convert these mean T scores for each age back into the corresponding number of questions correct by the use of the first and second columns. Thus, 41 (the mean for ten-year-olds), when located in the second column, is found to correspond to an original score of 17.

14. Make a diagram similar to that shown in Fig. 34, placing the first row of ages on the horizontal axis, and the last row of corresponding original scores on the vertical axis. Thus 8.5 is plotted with 1.8, 10.5 with 17, and similarly for the other pairs.

15. Smooth the diagram and prepare a table for converting every possible original score into its appropriate age score.

CHAPTER XXXIV
CONSTRUCTION OF SCALED SCORING
INSTRUMENTS

Derivation of Product Scales.—*Hillegas' English Composition Scale* and *Thorndike's Drawing Scale* are typical instances of product scales. Most of them are constructed as follows:

1. The scale constructor selects many specimens of, say, composition which vary by small amounts from compositions of zero merit up to, say, the highest quality of composition produced by the best authors.

2. He asks many presumably competent judges to arrange the compositions in order of merit and also to designate the specimen which is, in their judgment, of just zero merit.

3. He computes from these rankings the per cent of judges who rated specimen A better than specimen B, better than specimen C, and so on. Then he computes the per cent of judges who rated specimen B better than specimen C, better than specimen D, and so on. He continues this process until he has a table showing the per cent of judges who rate each specimen better than every other specimen. The per cent of judges rating a very poor specimen better than a very good one is likely to be zero, while the per cent rating the specimen of high merit better than the specimen of low merit is likely to be 100. Per cents of better judgments will range all the way from zero to 100.

4. He subtracts 50 per cent from all the above per cents.

5. He determines the P.E. (see Chapter XXXV) difference in merit between each specimen and every other specimen by looking up these remainder per cents in Table 49. Table 50 illustrates the process.

6. He not only determines the P.E. difference AB, AC, AD, etc., and BC, BD, BE, etc., directly, but he determines these differences in many indirect ways as well. Thus, for example, the distance NA, above, equals TN minus TA, the distance BN equals AB minus AN, the distance LE equals [(TL) minus (TA + AN + NB + BK + KE)]. There are many other indirect ways of determining the P.E. difference between any two specimens.

7. The mean of all possible direct and indirect determinations of P.E. differences is computed to get the true difference. The greater the indirectness the less the weight given to the determination in computing this mean P.E. difference.

8. He arranges the specimens in order of merit recording the P.E. distance each is above the preceding one, thus:

<i>Specimen</i>	T	A	N	B	K	E	L
P.E. Distance.....	0	1.0	1.0	1.5	.55	2.0	

9. He records from the original data the number of judges indicating each specimen as of just zero merit. Some will indi-

TABLE 50

DETERMINING P.E. DISTANCES IN MERIT BETWEEN COMPOSITION SPECIMENS

SPECIMENS	A > T	N > A	B > N	K > B	E > K	L > E
Per Cent.....	50	75	75	84.41	64.47	91.13
Per Cent Minus 50	00	25	25	34.41	14.47	41.13
P.E. Difference...	00	1.0	1.0	1.5	.55	2.00

cate, say, K, some B, some N, some A, some T, and some specimens which are below A and T in merit.

10. He computes the median zero specimen. Let us suppose that the median specimen is found to be A.

11. He computes the P.E. distance each specimen is above the zero specimen and calls this its scale value. Since A and T are of equal merit the scale becomes as follows:

<i>Specimen</i>	A or T	N	B	K	E	L
Scale Value.....	0	1.0	2.0	3.5	4.05	6.05

12. Beginning with the zero specimen he selects others above it such that distances between specimens will be about 1 P.E. Smaller scale steps are probably not desirable for scales which are to be widely used because a difference of 1 P.E. is a difference which only 75 out of 100 judges can see. Smaller-step scales may be valuable for scientific work, or for use by individuals who are specially expert in detecting subtle differences in merit. When two or more specimens have approximately the same scale value they may all be presented in order to give a wider range of composition type, or that one may be selected which shows the least disagreement among judges. The *Thorndike Extension of the Hillegas Scale* adopted the former method and the *Nassau Extension of the Hillegas Scale* the latter.

Validity and Constancy of Judgment Units.—What is the validity of this P.E. as a unit of measurement? Product scales were made possible by the formulation of the now famous Cattell-Fullerton theorem, and by the ingenious application by Thorndike of this theorem in the construction of educational scales. Courtis has reported an experiment which was conducted to test the validity of this basic theorem; namely, *differences which are equally often noticed are equal unless they are always noticed or never noticed*. Courtis wanted to know whether differences which are equally often noticed really are equal. To test this he made a product scale of areas instead of compositions or specimens of handwriting. After determining the differences between areas of variously shaped figures by means of judgments, he determined the differences by actual measurement. The differences as determined by judgments followed the principle of Weber's law, i.e., when the area was small, a slight increase or decrease in area could be seen; when the area was large, a considerable change of area was necessary in order that judges might be able to notice the difference. In other words, equally often noticed differences were equal for areas of about the same size only. The theorem does not hold for widely separated areas. Does it hold for specimens of penmanship widely separated in merit? Presumably it does not, if there are absolute differences in merit of handwriting in the same sense that there are absolute differences in area.

Even if this last is true, we need not lose confidence in our product scales. Education is interested in many kinds of differences. It would be valuable to know them all. There are absolute differences such as Courtis points out. There are difficulty differences, and this is the kind of difference Woody's arithmetic scales bring out. Product scales measure judgment differences. The values on percentile, age, and grade scales are determined by how difficult pupils actually find the test elements. These scales could be converted into product scales by determining the difficulty of each test element, not by the achievement of the pupils, but by the opinion of adults. This has not often been done simply because education is far more concerned with how difficult test elements actually are than how difficult somebody thinks they are. But in the realm of composition, handwriting, and the like, we are not primarily concerned

with difficulty but with merit, and we are less concerned with an absolute merit than we are with the merit as determined by the opinion of competent judges, in the way that competent judges practically operate outside or inside the schools.

Is the judgment scoring unit constant? The meter was originally defined as one ten-millionth of the distance from the pole to the equator. Alteration of this distance through the centuries due to the contraction or expansion of the earth would, of course, alter the meter, especially if a redetermination became necessary because of the loss of the meter bar carefully preserved at Paris. Alteration of this distance due to the subjectivity of the determiner would also alter the meter. As a matter of fact no two determinations of the pole to equator distance have turned out to be exactly the same. Consequently the meter is now measured in terms of so many wave lengths of a certain radiation. What forces are operating to produce an inconstancy of P.E. in, let us say, a composition scale?

Only the two most likely forces need be mentioned. First, it is possible to discriminate finer shades of composition merit. There is certainly room for improvement in this respect. So far as most of us are concerned there is "low visibility" when it comes to evaluating composition merit. The effect of a more microscopic eye would be to make P.E. smaller than it is at present. Second, it is possible that future judges will have a different opinion from present-day judges as to what constitutes merit in a composition. It is conceivable, but scarcely probable, that a literary dictator will arise whose popularity will be so great as to completely change the current of the world's literary appreciation. The nibbling of literary radicals is undoubtedly producing small but continuous changes in the weight we attach to each of the numerous factors entering into a composition.

Peculiarity of Product Scales.—Composition, handwriting, and drawing scales are peculiar in that they are not tests at all. They are scoring instruments. For this reason as well as for the manner of their construction they are called *product scales* to contrast them with percentile, age, T, and grade scales which together are usually called *performance scales*. Collection of the pupils' composition specimens is the composition test. The composition scale is only the scoring instrument. In the case of performance scales the dramatic instrument is not the scoring

instrument but the testing instrument. The following table will make clearer the relation between what is scored, the scale, the scoring instrument, and the scale unit:

<i>Thing Scored</i>	<i>Scale</i>	<i>Scoring Instrument</i>	<i>Scale Unit</i>
Man's height	Distance	Yard stick	Yd. ft. in.
Until train leaves	Time	Watch	Hr. min. sec.
Heat of water	Temperature	Thermometer	Degree
Courtis Arith., Series B			
(a) Speed	Speed	None	The example
(b) Accuracy	Accuracy	Correct answers	The example
Woody Arith., Series B	Difficulty	Correct answers	P.E.
Thorndike-McCall			
Reading Scale	Difficulty	Correct answers	G, T, or age
Starch Handwriting	Quality	Handwriting specimens	P.E.
Nassau Composition	Quality	Composition specimens	P.E.

CHAPTER XXXV

STATISTICAL METHODS

1. RELATIONSHIP MEASURES

What is Correlation?—The idea of correlation is so familiar that it is found in literary masterpieces and in the fables of the street. This is especially the case with inverse or negative correlation. “For every grain of wit there is a grain of folly.” “The vulnerable heel of Achilles.” “The leaf spot of Siegfried.” “Beauty *vs.* Brains.” “Eye-minded *vs.* ear-minded.” “Idea thinkers *vs.* thing thinkers.”

Thus correlation is a method for determining *the correspondence and proportionality between two series of scores or measures for the same pupils*, or the same schools, or the same cities, or any other entity. When the correspondence is perfect and positive the coefficient of correlation (r) is $+1.0$, when it is perfect, but negative, r is -1.0 . Correlation is *positive* when one series of scores tends to increase as the other increases, and *negative* when one tends to increase as the other decreases. A coefficient of correlation may be any size from $+1.0$ through 0 to -1.0 .

<i>Pupil</i>	<i>Test I Score</i>	<i>Test II Score</i>	<i>Test I Score</i>	<i>Test III Score</i>	<i>Test I Score</i>	<i>Test IV Score</i>	<i>Test I Score</i>	<i>Test V Score</i>
A	2	6	2	12	2	6	2	12
B	3	8	3	10	3	10	3	8
C	4	10	4	8	4	8	4	10
D	5	12	5	6	5	12	5	6
	$r = +1.0$		$r = -1.0$		$r = +.8$		$r = -.8$	

Some Uses of Correlation.—Here are some of the questions which education often asks and correlation can answer: How reliable is this mental or educational test? Does increasing its length or repeating it increase its reliability? Do these two tests measure the same aspect of reading ability, as they claim? Which one of a group of tests is most representative of all of them? Is there any justification for the popular assumption that pupils who are best in English tend to be poor in mathematics?

Do those who work most rapidly in arithmetic tend to work most accurately? How reliable is a teacher's examination in history? How close is the agreement between a test and a teacher's judgment? How close is the agreement between school marks and success in life? These and hundreds of other such questions involving a relationship between two series of measures can be answered by correlation.

Here are a few statements that correlation does not permit: When correlation is .8, 80 per cent of the pupils show perfect correspondence. When correlation is positive but less than perfect a larger score in one series *always* accompanies a larger score in the other series. When there is a high correlation between two series of facts one has caused the other, or correlation implies causal relation.

How to Compute Correlation by the Standard Method.—There are several formulae for the computation of r . The standard formula when the relationship is approximately rectilinear is Pearson's product-moment formula, which may be written thus:

$$r = \frac{\frac{Sfxy}{N} - (cx)(cy)}{\sqrt{\frac{Sfx^2}{N} - (cx)^2} \sqrt{\frac{Sfy^2}{N} - (cy)^2}}$$

Most educational relationships are rectilinear or are sufficiently so to make it permissible to employ the product-moment formula. But it is well to construct and inspect a scatter diagram (see Fig. 35) to determine whether the general drift of the diagram is rectilinear or curvilinear. If it is pronouncedly curvilinear the investigator is referred to some complete text on statistical methods for the appropriate formula.

Figure 35 shows in one diagram two sample scatter diagrams for two groups of twenty-five children. The circles show the relationship between attendance and distance. Each circle indicates one child's attendance record and distance from school. The general drift of the relationship is a straight-line or rectilinear drift. The crosses show the relationship between attendance and distance for twenty-five other pupils. Remember that the diagram is merely for illustrative purposes. It is extremely improbable that one group of pupils (circles) would show a

THE CIRCLES SHOW AN APPROXIMATELY RECTILINEAR RELATIONSHIP. THE
CROSSES SHOW A CURVILINEAR RELATIONSHIP.

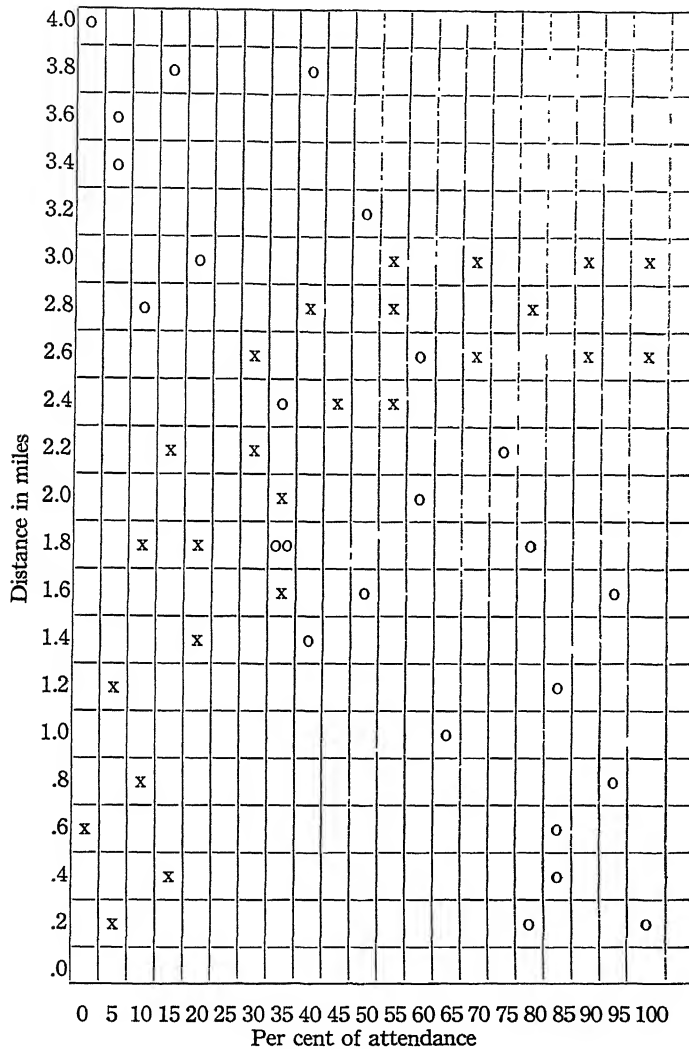


FIG. 35

decided negative correlation and another group (crosses) a decided positive correlation. But the important point to note

about the diagram is that the circles show a rectilinear drift whereas the crosses show a curvilinear drift.

The procedure for computing r is given in Table 51. Such a contingency table may be used not only as a starting point for computing a product-moment coefficient of correlation, but it also makes unnecessary the construction of a scatter diagram, such as Fig. 35. Inspection of the contingency table will show whether the relationship is sufficiently rectilinear to make the product-moment method applicable.

Table 51 is read thus: There were 3 pupils who lived between 3.4 and 4.0 (inclusive) miles distant from school whose per cent of attendance was between 0 and 10 inclusive, and similarly for the remainder of the contingency table.

There is no particular virtue in grouping the per cents in step-intervals of 15, or the miles in step-intervals of 0.8. The per cents could be grouped in step-intervals of 5, 10, 15, or any amount that is convenient. Likewise, the miles could be grouped in step-intervals of 0.2, 0.4, 0.6, 0.8, or any amount that is convenient. The size of the step-intervals chosen for Table 51 gives 7 steps for attendance, and 5 steps for distance. As a rule it is better to have a step-interval of such size as to produce not less than 10 nor more than 20 steps in each of the two items. The steps are made fewer in Table 51 so as to simplify the presentation of the correlation procedure.

The steps in the process of computing a coefficient of correlation from a contingency table follow. (1) Construct contingency table. (2) The total frequencies in the first column are 4. The total frequencies in the second column are 2, and so on for the other columns. The grand total of frequencies is 25. (3) The total frequencies for the first row are 5, for the second row, 4, and so on. The grand total of frequencies is 25, thus checking the preceding determination. (4) The assumed mean for attendance is 50, as shown by the vertical double ruling. The assumed mean for distance is 2.1, as shown by the horizontal double ruling. Other assumed means might have been taken, though assumed means near the center of each frequency distribution are more convenient. (5) The step-deviations from the assumed mean for attendance are shown in the x row. The step-deviations from the assumed mean for distance appear in the y column. (6) The product of each x multiplied by its corresponding

TABLE 51

SHOWS HOW TO COMPUTE A COEFFICIENT OF CORRELATION WHEN DATA HAVE BEEN TABULATED IN A CONTINGENCY TABLE
(AFTER H. L. RIETZ)

DISTANCE IN MILES	PERCENT OF ABUNDANCE											
	0	15	30	45	60	75	90	f	y	fy	fy ²	fx
	10	25	40	55	70	85	100					+
3.4 to 4.0	-18	-4	-2					5	2	10	20	24
2.6 to 3.2	3	1	1					4	1	4	4	4
1.8 to 2.4	-3	-2		0	1			6	0	0	0	
	1	1		1	1	0						
1.0 to 1.6						2	-3	5	-1	-5	5	5
0.2 to 0.8						1	1	5				
f	4	2	5	2	3	6	3	5	-2	-10	20	24
x	-3	-2	-1	0	1	2	3	25		-1	49	57
fx	-12	-4	-5	0	3	12	9	3				
fx ²	36	8	5	0	3	24	27	103				
$cx = \frac{fx}{N} = \frac{3}{25} = 0.12 \quad cy = \frac{fy}{N} = \frac{-1}{25} = -0.04$												
$Sfx^2 = 103 \quad Sfy^2 = 49 \quad Sfxy = 0 - 57 = -57$												
$r = \frac{\frac{Sfxy}{N} - (cx)(cy)}{\sqrt{\frac{Sfx^2 - (cx)^2}{N} \sqrt{\frac{Sfy^2 - (cy)^2}{N}}}} = \frac{\frac{-57}{25} - (-.12)(-.04)}{\sqrt{\frac{103 - (.12)^2}{25} \sqrt{\frac{49 - (-.04)^2}{25}}}} = \frac{-2.2752}{2.8356} = -.80 +$												

f appears in the fx row. The algebraic total of the fx's is shown at the end of the fx row. $Sfx = 3$. (7) The product of each y multiplied by its corresponding f appears in the fy column. The algebraic sum of the fy's is shown at the bottom of the fy column. $Sfy = -1$. (8) The product of each x^2 multiplied by its corresponding f appears in the fx^2 column. $Sfx^2 = 103$. (9) The product of each y^2 multiplied by its corresponding f appears in the fy^2 column. $Sfy^2 = 49$. (10) The f in the first square in the first column and first row is 3. The x at the bottom of this column is -3 . The y at the end of this row is 2. The product of $(3) \times (-3) \times (2)$ is -18 , which is written in the upper right corner of this first square. The f in the second square of the first column is 1. The x at the bottom of this column is -3 , and y at the end of this row is 1. The product of $(1) \times (-3) \times (1)$ is -3 , which is written in the upper right corner of the square in question. The f in the third square of the third column is 3. The x is -1 , and the y is 0. The product of $(3) \times (-1) \times (0)$ is written in the upper right corner. The f in the last square of the last row is 2. The x is 3 and the y is -2 . The product of $(2) \times (3) \times (-2)$ is written in the upper right corner of this square. The other f's times the xy products are computed similarly. (11) The sum of the fxy products in the first row, i.e., the sum of -18 , -4 , and -2 is -24 . This sum is written in the fxy column in the minus sub-column. Were this sum positive instead of negative, it would be written in the positive sub-column. In like manner, the sum of the fxy products for each row is computed and written in the last column. Positive $Sfxy = 0$. Negative $Sfxy = 57$. (12) The cx is computed; $cx = 0.12$. (13) The cy is computed; $cy = -0.04$. (14) $Sfx^2 = 103$. $Sfy^2 = 49$. $Sfxy = 0 - 57 = -57$. (15) The values previously computed are substituted in the correlation formula shown at the bottom of the table. By solving the formula, r is found to be $-.80 +$.

By substituting age-grade scores for distance scores in Table 51, and by recomputing, the r for attendance with age-grade relation can be determined. In similar manner, the r between attendance and any other factor, can be computed. The first row of Table 52 shows the coefficients of correlation between attendance and each of the six factors as computed by Reavis.¹

¹ Reavis, George H., *Factors Controlling Attendance in Rural Schools*, Bureau of Publications, Teachers College, Columbia University, New York, 1922.

Additional rows show the correlation between each factor and every other factor.

For our present purpose the first row of Table 52 is the most significant. It tells us that those whose attendance record are excellent tend to live near the school to the extent of .45, tend to progress rapidly through the grades to the extent of .50, tend to make high marks in school to the extent of .33, tend to have good

TABLE 52

SHOWING THE COEFFICIENTS OF CORRELATION BETWEEN ATTENDANCE AND EACH OF SIX HYPOTHETICAL CAUSES OF ATTENDANCE, TOGETHER WITH THE CORRELATION BETWEEN EACH CAUSE AND EVERY OTHER CAUSE
(ADAPTED FROM REAVIS)

CAUSES	2 DISTANCE	3 AGE GRADE	4 QUALITY OF WORK	5 TEACHER	6 SCHOOL PLANT	7 COM- MUNITY
1. Attendance .	-.45	.50	.33	.16	.07	.30
2. Distance . . .		-.20	-.13	-.10	-.06	.02
3. Age Grade24	.01	.08	.08
4. Quality of Work				.00	.08	.03
5. Teacher					.25	.35
6. School Plant						.17

teachers to the extent of .16, tend to have an excellent school plant to the extent of .07, and tend to live in a highly-rated community to the extent of .30. So far as these coefficients go, attendance appears to be most closely associated with age-grade relationship and distance.

How to Interpret a Correlation Coefficient.—Is an r of .30 or .37, according to the formula used, “high” or “low”? With r ’s as with intelligence, or wealth, or beauty, the customary criterion is that of relativity. There seems to be a sort of rough agreement among workers in this field that when r is

- 0 to $\pm .4$ correlation is low, or
- $\pm .4$ to $\pm .7$ correlation is substantial, or
- $\pm .7$ to ± 1.0 correlation is high.

There is, however, a more satisfactory way to interpret coefficients of correlation. When we have perfect correlation between two traits it is possible to predict accurately an

individual's position in one of these traits from a knowledge of his position in the other. As the coefficient of correlation goes toward zero such predictions become more and more uncertain. When the coefficient is exactly zero a prediction has no more accuracy than a sheer guess or a purely chance estimate. Kelley has worked out the data of Table 53. According to this table, when $r = 0$ the error of prediction is 1.00, where 1.0 is defined as a sheer guess. When $r = .1$ the error has been reduced to .995. The coefficient of correlation must be about .85 before the error is half-way between a guess and perfect prediction. Slight increases in the size of the coefficient above this point cause a rapid decrease in the error of prediction.

TABLE 53

SHOWS DECREASES IN THE ERROR OF PREDICTION FROM 1.00 TOWARD ZERO WITH INCREASES IN r FROM ZERO TOWARD 1.0, WHERE AN ERROR OF 1.00 IS A SHEER GUESS AND AN r OF 1.00 IS PERFECT CORRELATION

r	ERROR
.00	1.000
.10	.995
.20	.9798
.30	.9539
.40	.9165
.50	.8660
.60	.8000
.70	.7141
.80	.6000
.85	.5268
.90	.4359
.95	.3122
.97	.2431
.99	.1411

2. MEASURES OF TEST RELIABILITY

Self-Correlation Coefficients.—Self-correlation is the correlation between two duplicate tests given to the same pupils. Its chief function is to show whether one test is a sufficiently accurate measure of each pupil. Reliability is one criterion for evaluating a test. Self-correlation is one statistical technique whereby a test's reliability may be determined. If the self-correlation between two duplicate tests is 1.0, then one test is an absolutely accurate measure of each pupil in the trait which the test measures. This ideal is of course never attained.

How high should self-correlation be? No absolute standard can be given that will fit every situation. Where test results are used to commit children to institutions or to exclude them from important social or educational opportunities and the like, or where results are to be used for close theoretical reasoning self-correlation should certainly be above .9. But such a criterion is too drastic for most practical purposes, since the range of self-correlation for most standard tests is about .7 to about .9, while the range for typical teachers' examinations is much lower. A criterion of .9 or above would disqualify most educational tests and forbid as a public nuisance a professor's examination. Clara Chassell has found that the self-correlation of the marks of college professors on students who were rated through four full years is only .80! If the coefficient is not satisfactorily high it is evidence that one of two things needs to be done: (a) The test must be lengthened. How much it must be lengthened can be determined by computing the new correlation between the lengthened test and a duplicate of it. (b) If the test is not lengthened or not lengthened enough it must be repeated. How many times to repeat can be determined empirically by giving a test and its duplicate twice each and correlating the two series of averages, and if that is not enough, by giving each test three times and correlating averages, etc.

Prophecy Formula.—But this empirical process is very expensive in time, since twice as many tests as are needed must be given before it can be determined just how many are needed. The use of the Spearman-Brown prophecy formula will save half of this time.

$$r_x = \frac{Nr_1}{1 + (N - 1)r_1}$$

If the self-correlation of one test with a duplicate (r_1) is .8, and the information sought is how many times (N) the test must be given to yield a desired coefficient (r_x) of .9, substitute as follows and solve for N :

$$9 = \frac{N(.8)}{1 + (N - 1).8}$$

$$N = 2.25 \text{ times}$$

If the information sought is the r_x which would result from giving the same test or similar tests four times, substitute as follows and solve for r_x :

$$r_x = \frac{4(.8)}{1 + (4 - 1).8} = .941$$

Suppose that r_1 or .8 were the self-correlation between the average of two duplicate tests and the average of two other similar tests. In that case the N required to yield a self-correlation of .9 would be 2.25×2 or 4.5. The second formula would be interpreted as follows: 4 pairs of tests or 8 duplicate tests in all will yield an r_1 of .941.

Sometimes two equivalent forms are not available for determining a self-correlation coefficient. In this case one form may be administered and the total score made by each pupil on the odd-numbered items may be paired with that same pupil's total score on the even-numbered items. The coefficient of correlation between these two sets of scores gives the self-correlation for half of the test. If it is .6, the self- r for the whole test may be determined thus:

$$r_x = \frac{2 \times .6}{1 + (2 - 1).6} = .75$$

Index of Reliability.—But actually the whole test is more reliable, i.e., accurate than an r of .75 suggests. That r shows how close a test which is somewhat inaccurate corresponds to another test which is also somewhat inaccurate. Any test's correspondence with a perfectly accurate test, called its *index of reliability*, is shown by the formula:

Index of reliability = $\sqrt{\text{Obtained self-}r}$

Substituting, we have

Index of reliability = $\sqrt{.75} = .87$

P.E.score.—Since the index of reliability, like all coefficients of correlation whether self- r 's or inter- r 's, alters in size according to the range or variability in the group of pupils on whom it is based, there is needed a measure of a test's reliability which is free from this influence. P.E.score is such a measure. Its calculation and interpretation are shown in Chapter XV, F.

3. VARIABILITY MEASURES

Standard deviation.—In the preceding paragraph and frequently throughout this book there have been references to the variability in a group of pupils. The standard deviation (S.D.)

is one of the most commonly-used indices of variability. Its calculation is illustrated in Table 51. Thus:

$$\text{S.D. in attendance} = \text{Size of step-interval} \sqrt{\frac{Sfx^2}{N} - (cx)^2}$$

$$\text{S.D. in attendance} = 15 \sqrt{\frac{103}{25} - (.12)^2} = 37.05$$

Had the per cents of attendance been more closely bunched, the S.D. would have been less.

$$\text{S.D. in distance} = \text{size of step-interval} \sqrt{\frac{Sfy^2}{N} - (cy)^2}$$

$$\text{S.D. in distance} = 0.8 \sqrt{\frac{49}{25} - (-0.04)^2} = 1.12$$

4. AVERAGE MEASURES

Mean.—In Table 51, the mean per cent of attendance is given by the formula:

$$\text{Mean} = \text{Assumed mean} + cx \text{ (size of step-interval)}$$

The assumed mean is always taken as at the midpoint of the step-interval, hence

$$\text{Mean} = 52.5 + (.12 \times 15) = 54.3$$

The mean distance is computed thus:

$$\text{Mean} = 2.2 - (.04 \times .8) = 2.168$$

Median.—The median is now rarely used in connection with tests, the mean being generally preferred. In Table 51, the median per cent of attendance is 56.25, computed thus: $N = 25$. $N \div 2 = 25 \div 2 = 12.5$. Counting to the right along the f row to get a sum of 12.5, we have 4 + 2 is 6 + 5 is 11 and 1.5 of the 2. $1.5 \div 2 = .75$, which, multiplied by the step-interval of 15, gives 11.25. Then 11.25 added to 45, which is the beginning per cent of the step-interval in which the f of 2 falls, gives the median 56.25.

The median distance is 2.26, computed thus: $N = 25$. $N \div 2 = 12.5$. Counting down the f column to get a sum of 12.5, we have 5 + 4 is 9 and 3.5 of the 6. $3.5 \div 6$ is .583. $.583 \times$ the step-interval of 0.8 is 0.46. $0.46 + 1.8$, which is the beginning distance for the step-interval in which the f of 6 falls, gives the median, namely 2.26.

5. RELIABILITY OF r , S.D., MEAN, AND MEDIAN

The reliability of r and of the r for Table 51 are given by these formulae:

$$\text{P.E.}r = \frac{.6745(1 - r^2)}{\sqrt{N}} = \text{P.E.}r = \frac{.6745 \times .80}{\sqrt{25}} = .11$$

The interpretation of P.E.score in Chapter XV, F, shows how to interpret P.E. r and the subsequent measures of reliability.

The reliability of S.D. and the S.D. in attendance for Table 51 is given by these formulae:

$$\text{P.E.}s.d. = \frac{.6745 \text{ S.D.}}{\sqrt{2N}} = \text{P.E.}s.d. = \frac{.6745 \times 37.05}{\sqrt{2 \times 25}} = 3.53$$

The reliability of the mean and the mean attendance in Table 51 is given by these formulae:

$$\text{P.E.}mean = \frac{.6745 \times \text{S.D.}}{\sqrt{N}} = \frac{.6745 \times 37.05}{\sqrt{25}} = 5.0$$

The reliability of the median is $1\frac{1}{4}$ times the P.E.mean, and this is one reason why the mean is generally preferred to the median.

Those who desire to go more deeply into statistical methods may read the following books in order:

Garrett, Henry E., *Statistics in Psychology and Education*, Longmans Green and Co., New York, 1937.

Holzinger, Karl J., *Statistical Methods for Students in Education*, University of Chicago Press, Chicago, 1931.

Kelley, Truman L., *Statistical Methods*, The Macmillan Company, New York, 1923.

Thurstone, L. L., *Reliability and Validity of Tests*, Edward Brothers, Ann Arbor, Michigan, 1931.

Those who desire to go more deeply into measurement in research may read:

Good, Carter V., Barr, A. S., and Scates, Douglas E., *Methodology of Educational Research*, D. Appleton-Century Company, New York, 1936.

Monroe, Walter S. and Engelhart, Max D., *The Techniques of Educational Research*, Urbana, University of Illinois, 1938.

INDEX

- Abell, 420
 Abelson, 218ff.
 ability grouping, 161ff.
 achievement, 22ff., 156ff., 246ff.,
 284ff., 354ff., 405ff., 410, 463ff.
 actual classification, 191ff.
 Adams, 353
 administering tests, 137ff., 278, 281ff.,
 284, 286
 age score marks, 421ff., 502
 age scores, 65, 141ff., 153ff., 277,
 280ff., 282ff., 284ff., 290ff., 509ff.
 Alexander, 492
 all-year grouping, 170
 analogy, 50ff.
 aptitudes, 171
 assistance, 139
 attitudes, 39, 138, 271ff., 306ff.
 average, 529ff.
 axes, 480, 500
 Ayer, 420
 Ayres, 23ff., 42, 474
- B scores, 142, 356, 510ff.
 background, 228ff., 281ff., 405ff.
 Baker, 401
 bar diagram, 475ff.
 Barr, 530
 Barthelmess, 35, 168ff., 218ff.
 Batavia grouping, 171
 Bernreuter, 316
 Betzner, 343
 Bingham, 360, 401
 Bixler, 137ff., 419ff.
 Bovard, 299
 Briggs, 51
 brightness, 209, 225, 280, 354ff., 459,
 463
 Brinkmeier, 48
 Brinton, 474, 482, 492
 Brown, 527
 Bruner, 325ff., 335ff.
 Bryan, 413ff.
 Buckingham, 25
 Buros, 92
- Burton, 317
 Buswell, 399
- calibration, 52ff.
 captions, 482
 Carter, 31
 Caswell, 336
 Cattell, 24, 516
 Chaddock, 474
 chance, 74ff.
 Chapman, 52
 character, 160ff., 306ff., 444
 Chassell, 403ff., 527
 chronological age, 152, 212, 219, 354ff.,
 509ff.
 Clapp, 71
 classification, 156ff., 199ff., 289, 452ff.
 classification standards, 188ff.
 classification tables, 176ff., 204ff., 454ff.
 Cobb, 446
 combining scores, 146ff., 153ff., 203ff.,
 501ff.
 community, 228ff.
 completion, 37ff.
 composite, 146ff., 153ff., 203ff., 289,
 453ff.
 comprehensiveness, 41ff., 404, 407
 conservative classification, 186ff.,
 205ff., 458ff.
 contrast of opposites, 397ff.
 co-operative grouping, 172
 correlation, 51ff., 56ff., 218, 288ff.,
 519ff., 530
 Courtis, 25, 65, 69, 78, 379, 389, 499ff.,
 516
 Couzens, 299
 Coy, 466
 Crabbs, 367, 371, 411ff.
 criteria, 5ff., 147ff., 321ff., 423ff.
 crude scores, 140ff.
 cumulative records, 209, 446, 448,
 452ff.
 curriculum, 29ff., 157ff., 171ff., 211,
 267ff., 321ff., 406
 curve diagram, 475ff., 485ff., 502ff.

-
- Dalton grouping, 171
 democracy, 267ff.
 demotion, 176ff., 211, 289, 452ff.
 Denver grouping, 171
 departmental classification, 208
 Derryberry, 298ff.
 developmental history, 397
 diagnosis, 32ff., 42ff., 198, 226, 269ff.,
 283, 300ff., 354ff., 381, 383ff., 440ff.
 diagnostic tests, 398ff.
 difficulty, 62ff.
 divination, 218
 Downey, 316
 Dransfield, 368, 370
 dynamic tests, 299ff.
- efficiency, 361ff., 367ff., 402ff., 464
 effort, 356, 447
 Ellis, 421
 Elsbree, 317
 emotion, 307ff.
 empirical, 51
 Engelhart, 530
 Englehardt, 317
 environment, 11ff., 228ff.
 error, 56ff.
 essay, 37ff.
 Evenden, 317
 examinations, 39ff., 46ff., 74ff., 90,
 422ff., 435ff., 464
 examiners, 88
- F score, 356
 feeble-minded, 225
 Fernald, 399
 forms, 58ff.
 Franzen, 157, 160, 162, 165, 298ff.,
 413
 Freeman, 228
 French, 336
 frequency distribution marks, 419ff.
 frequency surface, 485, 487
 Fullerton, 516
- G age, 152
 G grade, 152
 Gans, 343
 Garrett, 530
 Gary grouping, 172
 Gates, 142, 147, 216ff., 371ff., 398ff.,
 400
 genius, 225ff., 359ff.
- gestalt, 11
 Good, 530
 grade score marks, 421ff.
 grade scores, 65ff., 141ff., 150ff., 277,
 280ff., 282ff., 284ff., 290ff., 416,
 502ff.
 graduation, 452ff., 460
 graphic methods, 473ff.
 Green, 69
 Greene, 401
 group tests, 72ff.
 grouping, 156ff., 289
 growth, 273ff., 356, 368ff., 404, 407,
 412ff., 459, 499ff.
 Gt, 199ff., 208ff., 289
 guidance, 356ff., 460ff.
- Haggerty, 147ff., 502ff.
 Hansen, 450
 happiness, 3ff., 33ff., 321ff.
 Harap, 336
 Hartshorne, 316
 Hawkes, 38, 90, 474
 health, 160, 298ff., 386ff.
 Henmon, 51
 heredity, 11ff.
 Herring, 3ff., 215ff., 278ff.
 Hildreth, 92
 Hill, 449ff.
 Hullegas, 25, 513, 515
 Hollingworth, 50ff.
 Holzinger, 530
 homogeneous grouping, 161ff., 209ff.
 Hopkins, 336
 Hull, 360
- identification, 139, 150, 292ff.
 individual tests, 72ff.
 instructions, 80ff., 139, 278ff.
 intangibles, 15ff., 20ff., 406
 integration, 309
 intelligence, 11ff., 153, 156ff., 161,
 209, 212, 215ff., 280ff., 354, 358ff.,
 387, 405, 410, 421ff., 430ff., 462ff.
 intensity, 308
 interests, 87ff., 165, 306ff., 386
 interpreting scores, 142ff., 280ff.,
 283ff., 285ff., 287, 354ff., 440ff.
 interview, 35ff., 38
 introspection, 387ff.
 irrelevancy, 43ff.
 isochron, 499ff., 502

-
- Jackson, 316
 James, 369
 John, 399
 Johnson, 72
 Jones, 65
 Jorgensen, 401

 Kelley, 90, 147, 316, 526, 530
 Kent, 65
 Kilpatrick, 415
 Kirby, 78, 379
 Kitson, 360
 Krey, 90
 Kruse, 52
 Kryzanowsky, 72

 late entrants, 208
 level, 209
 Lindquist, 38, 90, 420
 Lindsay, 168ff.
 linearity, 520ff.
 literacy, 14
 logarithmic scales, 478
 Lombardy, 316
 Long, 35
 Lorge, 360

 McComas, 50
 McGaughy, 165, 343
 McMurry, 21

 M score marks, 420ff.
 machine scoring, 72
 Maller, 92, 316
 Mann, 38
 marks, 419ff.
 matching, 37ff.
 materials, 317
 Mathews, 420
 May, 316
 mean, 148ff., 287ff., 409ff, 509ff.,
 529ff.
 mean deviation marks, 420
 mechanisms, 306ff.
 median, 514ff., 529ff.
 Meine, 52
 Meredith, 299
 miniature, 50, 83
 Modley, 492
 Monroe, 31, 42, 54, 395, 399, 413, 530
 Moore, 401
 Morrison, 142, 171ff.
 Morrissett, 323ff.

 Morton, 360
 Mossman, 343
 motivation, 87ff., 272ff., 306ff., 407
 multiple-choice, 37ff., 48, 77ff.
 multiple-response, 37

 norm, 61ff., 150ff., 279ff., 285ff.,
 288ff., 354ff., 364ff., 416, 503ff.
 Nygard, 413

 objectives, 3ff., 31ff., 342ff., 364ff.,
 406, 408ff.
 objectivity, 59ff.
 observations, 39, 228, 388ff., 479
 Odell, 413
 one-word, 37ff.
 oral tracing, 392ff.
 Otis, 69ff., 72, 323ff., 413

 Pearson, 35, 520
 percentage curve, 477, 483ff.
 percentage marks, 419
 percentile score, 65, 499, 502, 509
 personality, 311ff., 444
 personality quotient, 308ff.
 Peter, 413
 Peterson, 71ff.
 philosophy, 3ff., 29
 Pintner, 228, 413
 Pittman, 370ff.
 platoon grouping, 172
 practice tests, 368ff.
 preliminary tests, 81ff.
 Pressey, 69, 72
 probable error, 288ff., 441ff., 464ff.,
 514ff., 516, 528, 530
 processes, 342ff.
 product moment, 520ff.
 product scales, 513ff.
 prognosis, 171, 227, 356ff.
 projected Gi, 431ff.
 promotion, 176ff., 289, 452ff.
 prophecy formula, 527ff.
 purposes, 4ff., 37ff., 58, 299ff., 321ff.,
 342, 406

 questionnaire, 39
 questions, 139
 quotients, 209ff., 211ff., 225, 280, 356,
 357, 410, 413, 459, 463

 Rand, 413
 range, 220

-
- rate, 209, 459
 rating, 39, 311ff., 413ff., 440
 Reavis, 446, 524ff.
 record sheet, 150ff., 199ff., 278ff., 292ff., 433ff.
 Reeves, 170
 reference points, 495ff.
 reliability, 44, 55ff., 147, 218, 288ff., 464ff., 526ff., 528, 530
 remedial, 300ff., 364
 reports, 445ff.
 review term grouping, 170
 Rice, 23ff.
 Rietz, 521
 rivalry, 88ff.
 Roback, 316
 Roberts, 72
 Robertson, 446
 Robinson, 52
 Rogers, 51
 Rorschach, 316
 Rosanoff, 65
 Ruch, 48, 401
 Ruger, 68
 Rugg, 311
 Ruml, 52
 Russell, 399ff., 420ff.

 sampling, 39ff., 50, 218, 509ff.
 Sauvain, 162ff.
 scale, 62ff., 65
 Scates, 530
 school practices, 267ff., 286ff., 345ff.
 score cards, 317
 scoring, 66ff., 70ff., 278, 281ff., 284, 286ff., 435ff., 513ff.
 sectioned bar diagram, 485
 sectioning, 161ff., 209ff., 289, 459
 sector diagram, 484
 Segel, 469
 self grouping, 172
 short term, 210ff.
 significance, 42ff.
 Sims, 420
 Smith, 54
 Sobel, 444
 social, 158ff., 227ff.
 Somers, 33ff., 421
 Spearman, 527
 specific abilities, 161ff.
 Speer, 218
 Spence, 421, 466

 standard deviation, 498ff., 507ff., 528ff.
 standardization, 60, 140
 standards, 273, 283, 317, 404
 statistical, 42ff., 64ff.
 statistical classification, 176ff.
 Stenquist, 72
 Stoddard, 401
 Stone, 24
 Stratemeyer, 317
 Stratton, 51
 subject grouping, 170ff., 208
 subjectivity, 37ff.
 subtest, 285ff.
 survey tests, 382, 394ff., 415ff.
 Switzer, 170, 339ff.
 Symonds, 212, 316, 366, 401, 413, 421

 T score, 65, 356, 420, 497ff., 502, 505ff., 509ff.
 tabular methods, 492
 Teachers' Lesson Units, 337ff.
 Terman, 42, 161, 225, 357, 499
 Terry, 30
 test lessons, 368ff.
 tests, 68ff., 91ff., 215ff.
 thinking, 31
 Thomas, 401
 Thorndike, 15, 24, 51, 69, 81, 82, 141, 165, 216ff., 228, 360, 394ff., 415, 474, 499, 500, 513, 515
 Thurstone, 530
 time, 139, 483
 Toops, 52, 72
 Torgerson, 413
 total score, 174, 502
 Trabue, 25, 218
 transfer, 219ff.
 Traphagen, 401
 true-false, 37ff., 77ff.
 Tyler, 466

 Uhl, 393
 units, 497ff.
 urges, 306ff.

 validity, 29ff., 34ff., 36ff., 49ff., 218, 516ff.
 Van Wagenen, 500
 variability, 528ff.
 Vincent, 35
 Votaw, 420

-
- Walker, 416, 492
Weidemann, 48
weighting, 77, 146ff., 153ff., 289, 453ff.
Wells, 45ff., 401
Wilkins, 158
Winnetka grouping, 171
Witty, 158
Wood, 72, 446
Woodworth, 45ff.
Woody, 25, 42, 86, 141, 516
work sheet, 431ff.
Wrightstone, 310, 466
Wylie, 52
XYZ grouping, 161ff., 209ff., 289, 459
Young, 71
zero, 476ff., 483, 496ff., 515
Zubin, 76